10

Robust Institutional Design

What Makes Some Institutions More Adaptable and Resilient to Changes in Their Environment than Others?

Jenna Bednar

Abstract

Institutions are designed to alter human behavior. To remain effective over time, institutions need to adapt to changes in the environment or the society the institution is meant to regulate. Douglas North (1994) referred to this property as *adaptive efficiency* and suggested the need for a model of how institutions change to remain effective. This essay contributes to a theory of adaptive efficiency by relating it to the burgeoning literature in robust system design. It reviews five models of institutional change, paying particular attention to each model's ability to explain institutional adaptation. It isolates three common structural features of a robust, adaptive institutional system: diversity, modularity, and redundancy. It illustrates the theory with a brief application to federal systems, and closes by describing some open research questions relating to institutional adaptive efficiency.

Introduction

> It is adaptive rather than allocative efficiency which is the key to long-run growth. Successful political/economic systems have evolved flexible institutional structures that can survive the shocks and changes that are a part of successful evolution. But these systems have been a product of long gestation. We do not know how to create adaptive efficiency in the short run. —North (1994:367)

Institutions—formal rules, informal norms, and their means of enforcement—are incentive systems that shape and coordinate human behavior (North 1995, 2005). Institutions incentivize behavior by providing information, assigning roles to participants, and allocating resources and outputs. Formal institutions are designed to fit within a particular context, but the nature of the problem

can change. North's intuition was that an institution must be able to adapt to changing circumstances for it to maintain its functionality. This essay considers what we know about how to construct institutions to make them as effective as possible over time. The answer that I will propose does not focus, as one might think, on how to make institutions more consistent or sturdy, but instead on how to make them more pliable. I pay particular attention to advances in the theory of robust system design which puts scholars in a better position now to understand how institutions might be adaptively more efficient than when North wrote.

The aim of this essay is to introduce the reader to different theories of institutional dynamics. I've divided the essay into five sections. I begin with a review of how institutions are defined, explain the need for adaptation, and discuss the difference between two related concepts: robustness and resilience. In the second section I describe current theories of institutional change. In the third, these theories are applied to describe the system features that lead to robustness, paying particular attention to diversity, modularity, and redundancy. In the fourth section, I describe some of my own work on federal system robustness to illustrate the theory of robust design. In the final section I lay out a few (of the many) open questions regarding adaptive efficiency of institutions and robust system design.

Adaptive Efficiency as Robust Response to Changing Circumstances

In evaluating the performance of institutions—their ability to induce behavior that best meets socially defined goals—political economists tend to focus on how well they incorporate stakeholder interests or distribute burdens and benefits. Some examples include political economists like Boix (2003) and Acemoglu and Robinson (2012), who argue that political stability and economic growth follow from institutions that share decision making broadly and redistribute gains; North and Thomas (1973), who argue that property rights are key to growth because innovation is incentivized when it can result in private as well as social rewards; Weingast (1995), who in examining claims that federalism boosts markets, suggests that there are certain institutional conditions, in particular decentralization of fiscal authority, but centralization of rights provision and guarantee of free trade; and Ostrom (1990), who describes the institutional features that promote successful localized management of common pool resources through social enforcement rather than external, legal regimes.[1]

[1] Formal (written) and informal (social) institutions are not necessarily mutually exclusive or independent; to the contrary, in virtually every circumstance, formal institutions rely upon informal institutions, whether through socially based interpretation of the rule, monitoring, or enforcement.

While these studies focus on institutional features that increase allocative efficiency, an institution's long-term effectiveness depends on its capacity to adjust to changes in its environment. Examples abound of institutions rendered inefficient from changes to the nature of the problem they are designed to resolve: regulatory rules that cannot keep up with changes in technology; a federal system so symmetric and centralized that it cannot handle a rise in provincial nationalism; a foreign policy that is structured around nation-state relations and therefore has difficulty adjusting to terrorist threats from stateless groups. Institutions built to shape behavior at one state of the world—one specific environmental context—can fall far short in their effectiveness as circumstances change.

Douglas North (1990:80) introduced the phrase "adaptive efficiency" to describe the performance of an institution when its environment changes. This measure of performance shifts attention from the output of the institution—the specific allocation of distributable benefits—to its ability to adapt in order to be able to continue to fulfill its function. If the function of an institution is to allocate power or goods, its efficacy at performing this function is related to the fit between the behaviors generated by its incentive scheme and the environmental context where the behaviors operate. As the context changes, the fit between behaviors induced and context may slacken, decreasing the effectiveness of the institution.

North's sense of adaptive efficiency is closely related to *resilience*, a term used by evolutionary biologists to describe the capacity of a system to return to a given state after a perturbation shocks the system. Evolutionary biologists pay particular attention to consistency of the relationship between components in the system (Holling 1973). If the system fails to recover, it transitions into a new state. In an ecology, resilience would be a measure of how quickly some measure of the system, perhaps the population of a species or its distribution in the environment, returns to prior levels following an irregular disturbance such as a natural disaster.

Social systems are similar to ecologies: both have separate population groups with distinct interests interacting in an environment. In social systems, however, the interaction between agents is shaped by institutions, and institutions have properties that are different from the forces that motivate species behavior in an ecology. Importantly, institutions are both built and adaptive. They are designed by humans to fulfill a particular purpose. Institutions may evolve over time; they may also be redesigned or co-opted to fill new functions. To distinguish institutional analysis from the system features studied in biology, social scientists prefer the term *robustness*.

Like resilience, a robust system is one that maintains functionality despite perturbations (Jen 2005); it exhibits consistencies (Krakauer 2006) but is also capable of adaptation. Robustness shares resilience's concern for system recovery, but emphasizes system functionality and often relies, at least partially, on human-designed elements. Systems that may exhibit robustness include an

engine that propels a car, organs with the human body regulation of the financial system, or the management of depletable natural resources in a socioecological system. System robustness, then, may refer to how fail-safe the engine is, or how well the body handles disease (sometimes with medical intervention), or the capacity of the financial system to manage economic stress, or whether the population survives extended drought (see Carlson and Doyle 2002; Csete and Doyle 2002; Walker et al. 2004; Jen 2005; Anderies and Janssen 2013).

Importantly, neither resilience nor robustness is synonymous with *stability*. With stability, the emphasis is on non-oscillation. An institution remains consistent, its component parts unchanging, and its relationship to other institutions unvarying. Robustness, in contrast, recognizes that institutional variability is to be expected in a changing environment. To maintain functionality the institution may need to be transformed, either through evolution or intentional (re)design. A stable institution would uphold the same rule by the same methods, while a robust institution might adjust the rule or means of supporting it depending on the context, as long as the adjustment continued to lead to the same social goal.

With this discussion to tie North's concept of adaptive efficiency to theories of robustness, let us consider different explanations for how institutions might adapt to respond to change.

Five Models of Institutional Change

When problem landscapes change and make institutions less effective at promoting desired behavior, a model that captures how institutions drive behavior, and how they respond to their environment, can help analysts to understand the causes of institutional–environmental mismatch and how to repair it. This section builds on the work of the economist Allan Schmid (2004, chapter 13) to describe five models of institutional change. Systems—engineered, social, biological—vary in their adaptive capacities and design features. Engineered systems (e.g., a car's engine, software) are generally nonadaptive, so their robustness depends on physical recovery features, such as redundancy or regularized human review. Biological systems, on the other hand, are almost purely adaptive; exceptions like genetic modification only highlight the system's overwhelming reliance on evolutionary processes for change. The institutions that drive human behavior in social systems combine both designed and adaptive elements to respond to external and internal stresses. System dynamics are expressed in terms of immediate means of recovery as well as slow adaptation to long-term changes, and both evolutionary forces of adaptation and intentional tweaks can be modeled in considering how well institutions respond to a changing environment. Social scientists can learn from engineered systems theorists as well as from evolutionary biologists, but must adapt their models to suit the particular factors that characterize social systems.

Robustness theories, borrowing language from biology but retooling it for social processes, describe institutional adaptation through a combination of innovation (mutation), selection, and reproduction. Innovations may be intentionally created or accidental, akin to the mutation of a biological organism. Because of the element of human choice in institutional design, models of functionality and power are important to keep in mind when considering robustness. The institutions embedded within a written constitution, or the meaning of the constitution itself, may change either as a result of conscious human deliberation or of adaptive social processes.

Schmid (2004) identifies four fundamental theories of institutional change: functional, power, isomorphic (path dependence), and learning models. To Schmid's four types, I add one more: behavioral spillovers. The five model types are not analytically exclusive; North, for example, employs all in his institutional analysis. To understand the potential and limitations of current modeling endeavors, I review these modeling types here. The categorization is useful for our purpose of thinking about how an institution's structure affects its ability to respond to its environment.

Functional Theories

In a so-called functional theory (called that by Schmid; other terms which may be more familiar to social scientists are mechanism design, positive political theory, or new institutional economics), an institution plays the role of intermediary, translating action into outcome. The institution is chosen by a social choice function such as majority rule or unanimity. It does not change until another social choice process alters it, either through amendment or replacement. Institutions perform their functions according to a measure of efficiency (e.g., the transaction costs that increase the price of market operations). In functional theories, exogenous changes, such as a new technology, can change the efficiency of the existing institution. If new technologies make alternative institutions more efficient, the agents may adopt a new institutional form to improve social welfare. Schmid attributes the theory to new institutional economics, particularly in the work of economists such as North (1990), Williamson (1985), and Weingast (1995). The functional theory of change is essentially punctuated equilibrium: once an institution is adopted, it continues until an exogenous change makes another institution more efficient. Institutional change is a conscious and rational reaction to an exogenous change to the environment.

Theories of Power

Power theories differ from functional theories by focusing on the interests of those in control. Where functional theories focus on the maximization of utility

according to a social welfare function (e.g., wealth maximization), power theories remind us that institutions have distributive consequences. Those who are likely to benefit from a new institution will support it. When institutions are chosen by a subset of agents, they will favor the institution that maximizes their own welfare, rather than that of society as a whole (Knight 1992). Agenda setters will manipulate voting rules to ensure that their favored rule is selected, including, where appropriate, maintaining the status quo (Riker 1982). Those in power are unlikely to support institutional change unless it would benefit them even more than the current arrangement. Therefore, institutions can fail to respond to changes to their environment when doing so would worsen outcomes for the powerful elite.

Like functional theories, the power theory of institutional change relies entirely on human agency as the source of change, but it does not stress the exogenous shock to the same degree. That is, within power theories, if agents in control could change the institution to give them an even greater share of the system's resources, they would do so. Therefore, the pure-type of power theory is not an evolutionary model and does not incorporate any nonconscious adaptive processes. However, the nature of power itself is changing, and thus the structure of power theories may change as well. Traditionally, power has been conceived as being held by a governing leadership, in possession of political legitimacy or economic resources. Emerging threats suggest a new brand of power that is distributed in a network, and network models of power are gaining influence (Granovetter 1973; Castells 2011).

For example, terror is a distributed threat that breaks from the standard mold of power theories. People's fear is a strong motivator of behavior; those willing to be ruthless do not need to be at the top of a political or economic hierarchy to wield power. Terrorism is an institution in its own right to be explained; that is, the capacity of terrorist organizations to recruit, retain membership, organize themselves, and formulate, articulate, and achieve goals. It also poses an important challenge to institutions that do provide governance: a robust institution of governance, whether political or economic, should be able to withstand distributed threats as well as pointed ones.

Accordingly, theories of terrorism's power are moving into new analytical forms. For example, Caplan (2006) questions the efficacy of expected utility calculations of deterrence capacity when terrorists deviate from standard rationality assumptions. Galam and Mauger (2003) suggest how using a percolation model offers insights about how to fragment global terrorism threats into localized activities, which might then be countered through conventional means. Enders and Sandler (2011) describe the context sensitivity of terrorist organizations with arguments about feedback effects, cycles, and a threshold model affecting the likelihood of terrorist attacks. These models are more in line with theories of path dependence, to which we next turn.

Path Dependence

In institutional analysis, the most commonly encountered theory of institutional evolution is path dependence. The tendency for systems (biological and social) to exhibit path dependence is also one of the key problems that can limit institutional robustness. This form of adaptation is often maladaptive because the forces of change are tied to the institution itself rather than being responsive to the external environment.

The phrase "path dependence" is used interchangeably with "history matters," and often with just as much logical precision. In general, path dependence implies that the sequence (or as the theory is often applied, the set) of events in the past influences future outcomes, with an implication that path-dependent processes converge toward an equilibrium distribution of states, where new forms become less likely. But as Page (2006) has pointed out, path dependence is not a very precise term. At one extreme, where the institution is as twitchy as a colt to new sensory input, it changes all of the time. At the other end of the spectrum is extreme sensitivity to initial conditions, where the institution's shape is determined by the first inputs. With extreme sensitivity to initial conditions, the institution is dead to new experiences; it ignores environmental changes.

Path dependence captures diverse mechanisms including increasing returns, self-reinforcement, positive externalities, and lock-in (Page 2006). Schmid's (2004:262) description of path dependence focuses on the theory of *isomorphism*, an example of positive externalities. Due to bounded rationality, organizations will borrow routines, rules, and behaviors from other organizations rather than develop their own, despite mismatch (e.g., DiMaggio and Powell 1983). The most common model of the force that generates path dependence is increasing returns, where behaviors become less costly over time (e.g., as routines develop) and thus more likely (e.g., David 1985; Arthur 1994). Greif and Laitin (2004) build a model of self-reinforcement to describe a process of endogenous institutional change. Game play creates a feedback that changes the payoff structure, what they call a *quasi-parameter*.

Theories of self-reinforcement shed new light on power theories. Acemoglu, Ticchi, et al. (2011) explain the entrenchment of a corrupt system as follows: if elites can capture control to reduce redistribution, they can maintain this control through patronage. This is an example of Greif and Laitin's quasi-parameter: corruption makes the continuation of corrupt practices more likely, as proponents can be bought with the rents skimmed from corrupt practices. Similarly, Kollman (2013) relies on self-reinforcement to argue that federal systems will inevitably grow more centralized: as the federal government gains fiscal authority, its access to resources enables it to assume new powers. States become dependents of the federal government rather than competitive rivals.

Historical institutionalists examine institutional performance and change by situating institutions within a context of institutional and behavioral space,

studying how experience shapes decision making at particular moments. In the methodology of process tracing, scholars identify key causal mechanisms within the context of a specific case (e.g., Thelen 1999; Falleti and Lynch 2009). As Falleti and Lynch, quoting Goertz, put it: "Context plays a radically different role than that played by cause and effect; context does not cause X or Y but affects how they interact" (Falleti and Lynch 2009:1151; Goertz 1994:28).

Historical institutionalists are tapping into the influence of perception—belief systems—to study the way that people process information. North's intuition about how to build adaptive efficiency focused on the incorporation of information. When agents sense changes in their environment that create new opportunities, they shift institutions accordingly. These beliefs are path dependent (North 1995). Greif (2006:188) explains why belief systems constrain institutional forms: "institutional elements inherited from the past are the default in providing the micro-foundations of behavior in new situations." Human nature advantages traveling familiar paths. A society's historical experience with an institution, or components of it, should cause that society to implement familiar institutional components rather than ones that might appear to be more efficient, from a mechanism design perspective. There's efficiency in familiarity.

Path dependence's winnowing of possibilities may not even be time dependent. Bednar and Page (2012) ask: What if, with some small probability, a later game changes play in an earlier game? If this occurs, rather than opening up the possible paths, outcomes tend toward a single type. For example, in law, a belief revision would be a reinterpretation of the meaning of a prior Court decision. The Court adheres to precedence while innovating; if accepted, this new interpretation of the prior interpretation reduces the scope of legal arguments available to future arguments. Belief revisions may lead to greater coherence in the law or other institutions.

Path dependence, then, is a theory of limitations on institutional change. Dynamics lead toward convergence. To be sure, the advantages of path-dependent processes are significant. Institutions have structural integrity, are characterized by consistency, and are predictable. These features are key elements of a legitimate legal order. But sometimes what's best is to jump the tracks. Path-dependent processes don't allow that movement, by definition.

Learning and Evolution

Power models of change specified processes of selection, whereas path dependence added reproduction (perhaps too much). Both the power and path dependence models, however, have been curiously silent on mutation. In borrowing insights from their biological predecessors, evolutionary models of institutional change include mechanisms of mutation, selection, and reproduction. In the institutional realm, mutation is the introduction of a new idea (perhaps about

the structure of the institution), selection is the mechanism for choosing between ideas, and reproduction is the means of carrying adopted ideas forward across rounds of play.

In evolutionary models, institutions evolve through a process of *cumulative selection*. Mutation introduces a small change which is compared to the initial structure; the best of the two is kept and the process is repeated. Selection is not random or single-step, but preserves prior improvements. Genetic algorithms such as hill-climbing mechanisms feature this process. Note, importantly, that adaptive models require some measure of fitness that the two alternatives can be compared against. In an adaptive landscape, higher positions on "hills" represent improvements to utility.

Adaptation requires exploration of new ideas. Systems must learn more about their environment through experimentation or mutation. Too much experimentation or pursuit of new information can be inefficient when the system fails to use existing information. With adaptive efficiency, the system balances regularity with experimentation, whether the application is machine learning (Holland 1992; Holland et al. 1962), phenotypic consistency in biology (Fisher 1930; Krakauer 2006), or standard operating procedures in organizational culture, as captured, for example, in March's seminal article regarding the optimal trade-off between informational exploitation and exploration (March 1991). Given that a system cannot simultaneously exploit current best practice and conduct trials of new practices designed to reduce the error in existing practice, a robust adaptive system will have an internal regulator that allocates some energy to maintaining regularity while some subset of the system conducts trials.

Spillovers between Institutions

Although Schmid only described four types of models, there is a fifth type that recognizes the interaction between institutions and therefore the interdependence of their effects. One interesting output of the human genome project is our increasing awareness that selection does not work on a single gene. Genes are interactive in their effect on the phenotype's fitness. The same is true of institutions. In influencing agent behavior, the effects of institutional incentives *spill over* to contaminate or improve the effects of other institutions. One limitation of the game theoretic literature is that it almost exclusively looks at the functioning of a single institution. However, institutions exist in a fuller context—an ecology, if you will—of institutions. The institutional context creates a pattern of behaviors and beliefs; it creates a *culture*. This culture affects the performance of any one institution, and will affect its evolution. Analysis should take this context, the spillover effects between institutions, into account.

In the Bednar and Page models of culture and institutions, several institutions simultaneously influence behavior. Agents regularly carry over behaviors from one institution to another, even in the absence of any reward or other motivation to do so (e.g., Bednar and Page 2007). It is possible to predict the

direction that spillovers will flow based on the difficulty of optimally coordinating behavior with another agent (Bednar et al. 2012). A model of institutional path dependence (Bednar and Page, submitted) can lead toward an understanding of how the introduction of institutions might be optimally sequenced to avoid lock-in. In this model, institutional path dependence increases and then decreases in behavioral spillovers, weak punishment is important to maintain experimentation, early diversity of games is important, and, maybe most counterintuitively, to avoid lock-in, institutions should be sequenced to maintain the possibility of path dependence.

A model of institutional spillovers is complementary to the other four modeling types. Agents select behaviors to maximize their own return, incremental mutations in their behaviors help them to learn about the institutional environment, and behaviors can converge suboptimally, echoing the effect of the path-dependence models.

The possibility of institutional spillovers has implications for robust institutional design. With institutional interdependence, multi-institutional systems can have rapid transformations as mutations ripple across the system (see final section, A Few Open Questions, for design implications of rapid transformations as a direction for future research).

Characteristics of Robust Systems

Robust design implies the means to be both effective with current conditions (forceful) and flexible to respond to changed conditions (adaptive). Bearing in mind the lessons from the theories of institutional change, there is a set of desiderata for robust institutional design. Robust design implies overcoming the maladaptive tendencies of positive reinforcement while maintaining effectiveness. It implies reducing the negative aspects of power, including change in response to environmental circumstances rather than to new demands of an elite minority. It means not letting short-term concerns dominate long-term interests. Robust design is built to take advantage of beneficial adaptive forces. If possible, it will be mindful of spillovers between institutions.

Robust institutional design considers not the immediate problem of changing behavior, but focuses instead on the structure of the institution—or the system of institutions—so its performance is robust. The organizational form of the institution affects the institution's robustness. There are three characteristics—static features—of a robust system: diversity, modularity, and redundancy.

Diversity

Diversity plays a critical role in institutional system robustness (Page 2010); it is an engine for mutation and adaptation. In adaptive systems, the system

needs a means of mutation and a selection method to reproduce the beneficial mutations while allowing detrimental ones to die out. Designing an institution to prioritize diversity means to keep open the channels for input of diverse interests and new actions.

The robust system will seek diverse new information. The more diverse the experimentation, the more likely the system will encounter a modification that improves it. This insight was first articulated by the evolutionary theorist R. A. Fisher (1930:37), who noted that the "rate of increase in fitness of any organism at any time is equal to its genetic variance in fitness at that time." This claim, known as Fisher's fundamental theorem, suggests that biological organisms depend on genetic variation to survive complex environments. In social systems, the more a system is able to incorporate diverse ideas, the more likely it is to discover better solutions to problems (Page 2007, 2010).

Acemoglu and Robinson (2006) suggest that innovations are threatening to those in political control. They model the effect of the "political replacement effect" on innovation as a nonmonotonic function of the extent of political competition. Competitiveness varies from one extreme of noncompetitiveness, where elite power is secure, to highly competitive systems, where no elite can count on maintaining power. At these two extremes, elites will not block innovation. In systems where elites are powerful but feel threatened, however, elites will repress innovation to bolster their advantage as incumbents. Depending on the volatility in the environment, new ideas are unlikely to be sufficiently broad if they only represent the interests of the powerful. Given that asymmetric power distribution affects institutional change, a robust system's structure would reduce the influence of power, perhaps by fragmenting it.

Even without the obstacles to experimentation raised by those who are threatened by change, experimentation is costly for a social institution. As Greif (2006:191) writes, learning will be a "lengthy, costly, uncertain endeavor." A challenge for robust design is to sort out how to implement March's suggestion: to experiment while still capitalizing on existing knowledge.

Modularity

Modularity contributes to robustness in a number of ways: it breaks down the scope of the problem to manageable chunks, local diversity can be exploited appropriately, and failures can be contained within the module. Modularized systems can be self-similar or specialized. A self-similar modular system means, approximately, that the system's elements and properties are repeated at each scale. It can be fully horizontal, as in a distributed network (e.g., a terrorist network organized into cells). It can also include a vertical hierarchy: the U.S. political system has self-similar governmental structures at the federal, state, and local levels. A specialized modular system, in contrast, isolates components rather than integrating them fully. Car engines are modularized as is the U.S. government's separation of judicial, executive, and legislative powers.

When problems are chunked into modules, an otherwise overwhelming problem can become manageable. Travel and contact restrictions on Ebola patients minimize the spread of a highly contagious and deadly disease; care can be focused on contained areas rather than a worldwide pandemic. If the problem landscape varies at all, localization enables tailoring of solutions to meet particular needs.

Modularization can aid the challenge of harnessing diversity to explore new alternatives and be more responsive to a changing environment (Sanchez and Mahoney 1996). In a modularized system where each module mutates independently, new ideas can be explored without committing the entire system to the experiment. Each module can evolve on its own—or be redesigned by separate teams—and advances can be reintegrated into the whole. Baldwin and Clark (2000) credit the modularization of the computer industry with the industry's rapid innovation and growth; distinct design teams could specialize to develop each element.

In either case (self-similarity or specialized) if a module fails, whether a terrorist cell is knocked out or a liver fails, the remainder of the system is protected from the failure. Separation of governmental powers makes it possible for corruption charges to be contained to a single branch of government. Specialized systems require replacement of the failed part, but the failure does not spill over to the other units. Modularity of self-similar units fully buffers failure when the source is particular to the module; for example, if a terrorist cell is discovered, the remainder of the network remains operable.

Modularity can also compromise robustness. As with diversity, there can be too much of a good thing. If a system is too modularized, then beneficial change cannot diffuse through the system. Coordination throughout the system is impeded by modularity. The system is not as efficient.

In this respect, social systems have an advantage over biological systems. Recall the first two theories of institutional change described above; institutional change is not exclusively adaptive, but also involves human choice. In self-similar modularized systems, beneficial changes can be consciously copied by other cells. In federal systems, state governments watch the policy experiments of other states, imitating those which appear to work well. Renewable energy portfolio standards, recognition of gay marriages, and legalization of marijuana are all recent shifts in rules that have diffused across the states.

Redundancy

Institutional redundancy aids recovery from perturbations. Should one element of the system fail, a redundant pathway, with identical functionality, can play the same role. Successful parallel systems have two characteristics: (a) redundant components should have fully overlapping functionality and (b) as much as possible, they should have uncorrelated vulnerabilities. That is, while their capacity should overlap, they should fail for different reasons. Note that the

components need not be identically designed; actually, it is their diversity that makes them useful as insurance, because the redundant component can only be useful if it does not fail when the first does.[2] With redundant functionality, the system is more reliable than its parts (Bendor 1985; Bednar 2009).

At this point it is useful to remember what an institution functions to do: it is designed to change human behavior. It can do so by providing positive incentives, but it also often shapes behavior by threatening a punishment. This functionality introduces two kinds of considerations for redundant design: the problem of failing to punish undesirable behavior and the problem of punishing too much. With only a single trigger mechanism, the institution may fail and behavior not be redirected. The principle of redundant design would suggest two independent trigger mechanisms, but that risks the possibility that both will punish deviant behavior. Overpunishment can be as problematic as underpunishment.

This problem has a familiar cognate in statistics: the trade-off between the risks of making a Type I and Type II error. With the Type I error, or false positive, the null hypothesis is rejected incorrectly, whereas the Type II error, or false negative, fails to reject the null hypothesis when it should. As opposed to the reliability problem posed above, where the concern is that a component will fail to act—the Type II error—it is the opposite issue that statistics views as more problematic: rejecting a hypothesis incorrectly. As we know well from statistics, for any fixed sample size, Type I and Type II error risks are off-setting: to reduce the risk of one, you increase the risk of the other. The prescription is again straightforward. Consider the Type I problem first. Set the critical value (the threshold) based on an acceptable failure rate. To augment the power of the test (its ability to avoid Type II errors, false negatives), increase the sample size.

Translating the false positive to the federal context, the consequence of a safeguard failure that causes it to trigger a punishment too often could be huge if the safeguard triggers intergovernmental retaliation, including retaliatory opportunism and even withdrawal from the union. So from statistics we learn that these problems compete with one another, and if we fear convicting the innocent more than undue leniency, then we should solve that problem first. It would appear that adding redundant institutions will only augment the problem by making punishment more certain.

There is a theoretical solution. Until now we have conceived of redundancy too narrowly. There is a second form of redundancy that helps to solve the problem of overly frequent punishment. If an institution punishes too frequently, one problem may be that its observation of the agents' actions is biased. A second institution, with an independently drawn signal, could improve the efficiency of the system's reaction. Here the second signal (or institution) is not

---

[2] Biologists prefer the more accurate term *degeneracy* to capture the concept of redundant functionality, as in when the genetic code contains multiple codons for the same amino acid.

enforcement insurance, but confirmation that the action crossed the threshold of unacceptable behavior. The theory of the second signal is related to the results in Sah and Stiglitz (1985, 1986), who compare two organizational forms, the hierarchy and the polyarchy. In a polyarchy, a project is accepted if one (of two) agent accepts it. Both must reject it for it to be rejected. In a hierarchy, two agents are necessary to approve a project, and therefore only one is sufficient to reject it. The hierarchy echoes the reliability literature. It is a redundant system of safeguards, with one backing up the other (insurance); punishment is inflicted if either one of the safeguards is triggered. Likewise, equivalent to the polyarchy is a second form of redundancy; this time, two safeguards must be triggered before the punishment is levied. In this sense, the redundancy is confirmatory.

The optimal organizational form will depend upon the nature of the failure risk. As both kinds of failure—under- and overpunishment—exist in federal systems, in my work on federalism I describe how a system of institutional safeguards can play both insurance and confirmatory roles (for more detail, see Bednar 2009).

Federalism and Robustness

In any system to be studied, one must first identify the primary problem. In federalism, the animating problem area is the distribution of authority between the state and federal government. Although the "boundaries of federalism" are set constitutionally, governments regularly challenge these boundaries. It is the function of institutional safeguards to defend those boundaries—to ensure compliance with the Constitution. These institutional safeguards are diverse in composition, focus, and ability to punish transgressions. They are also flawed and often in disagreement as to the precise definition of the boundaries. With a theory of institutional complementarity, I define conditions where the safeguards improve one another's performance. With a systems view, minor disagreements between the safeguards about where to draw the boundaries (how to interpret the constitution) leads to the persistence of opportunism; for example, why states legalize marijuana (a federally regulated substance) and why such acts don't spell the beginning of the end of U.S. federalism, but instead, should be considered part of its normal operation.

My contribution is to consider the effect of rules as they sit in a context of other institutions, each acting simultaneously on individuals to influence social outcomes. Many scholars have examined one aspect of institutional interaction, as joint decision making (e.g., Montesquieu, Madison, or modern theorists, such as Scharpf, Tsebelis). I also examine how institutions can support one another functionally, building a theory of *institutional complementarity*. Each institution is imperfect, with enforcement gaps. Just as in complements, with institutions backing one another up functionally, compliance is improved

(although opportunism is never entirely eliminated, which can be beneficial, as discussed below).

The structural features of federal systems are ideally suited to meet both criteria of North's theory of adaptive efficiency. Experimentation is a useful way to explore the policy space, to determine whether any change to the distribution of authority might be welfare enhancing. Rather than a single government modifying its policy, if policy is decentralized then some governments may continue with established practice while others might experiment. Rather than conduct single experiments, the variety of states and municipalities guarantees that different policies will be tried. Subsidiarity—a principle of prioritizing decentralization—is the catalyst that boosts the likelihood of this experimentation: it is Brandeis's vision of the states as policy laboratories.

In the federal context, adaptation refers to changes to the distribution of authority. Adaptive systems require a method of mutation—an internal source of change—and a selection mechanism to separate the useful changes from the ineffective. In federal systems, governments continually push against the limits of their authority. This persistent low-level opportunism provides the system with a means to experience alternative formulations of federalism's boundaries. Minor challenges to the constitutional boundaries served as Brandeis's laboratories of democracy: the nation could learn whether there was improvement to the system. Opportunism—defiance of constitutional rules—is not only inevitable with a system of safeguards, it is key to the adaptive process, helping federal systems adapt to their environments.

A Few Open Questions

As I wrote this essay I kept a log of questions I had that to which I could not find answered in the literature (i.e., open questions). This collection is in no way exhaustive and is generated only by my own interests. Most are different takes on the same question: just how robust is a robust system? The interdependence of the system components leave it vulnerable to spectacular failure.

System Fragility

All systems balance robustness and fragility, something that could cause them to collapse (Carlson and Doyle 2002; Crutchfield 2009; Anderies and Janssen 2011). This fragility may be hidden; that is, it may not be understood until a perturbation exposes it. Complex systems can be highly interconnected, and actions can reverberate across the system producing large events (recall the usefulness of modularity). Examples of these sorts of phenomena include market crashes, mass extinctions, and power grid failures. Because there are internal capacity tensions, such as the trade-off between hyper-redundancy

and efficiency, robust systems have within them a fragility. They can fail. The question is: How great is their capacity to bear the perturbation, versus how efficient (often minimal energy) are they? As Janssen and Anderies (2007:43) put it: "The choice for society is not only whether to invest in becoming robust to a particular disturbance, but rather, what sort of disturbances to address and what set of associated vulnerabilities is it willing to accept as a necessary consequence."

Countering Maladaption and Leaping from Low Peaks

As discussed in the section on path dependence, institutional change can be maladaptive. If the process of change involves path dependence with feedback effects, structures can be reinforced even as they grow less compatible with the environment, in the sense of generating behavior that maximizes utility.

A less subversive form of maladaptation is a mechanism for exploration that only explores local alternatives; this system can get stuck on local optima, failing to make the transformation necessary to higher "peaks." A cumulative selection model is also vulnerable to getting stuck at suboptimal points. This may be *the* problem of adaptive efficiency. Although the system can be designed to explore locally, it is also designed to avoid big downturns, which means that it will fail to catch big opportunities if they require a complete rupture from current practice. Agents in this institutional system will also not be able to acquire information needed to learn that their ship is sinking, so to speak; their practices may be on the decline, but as long as it continues to be better than the immediate alternatives, the adaptive mechanisms will not help the community to endure.

Competition between Adaptive- and Present-Time Efficient Institutions

While robust institutions have long-term advantages, in many domains their acceptability is decided upon short-run performance. In quarterly stock performance reports or frequent elections, an adaptively efficient institution may not compare well against an institution that happens to work well for the immediate context. In the case of firms, the problem can be acute. Standout success is often rewarded with higher levels of discretion or larger budgets. The 1995 failure of Barings Bank, founded in 1762, was the result of actions by a single trader in Singapore who initially generated substantial profits. Similarly, one of the world's largest insurers, AIG, was brought to the brink of default (and rescued using public funds) because of the activities of a single division in London, which again was enormously profitable for a time (for detailed discussion of this example, see Sethi 2012; for a discussion on the "destabilizing effects of stability," see Minsky 1982).

Irreducibility

System theorists accept that the system may have properties that are distinct from its components. Vermeule (2011) phrases this lesson in terms of a warning: one mustn't be tempted to infer that because a system has particular characteristics—say, transparency, or democratic responsiveness—that each component of the system will have that characteristic. The canonical example of irreducibility is the existence of the unelected judiciary within the U.S. democratic system. It is the sort of contradiction that confounds the public and great scholars alike: an unelected, unrepresentative, unaccountable body of nine can block the intentions of the public's political agents. Still, the judiciary is heralded as one of the U.S. democratic system's three key pillars, with a popularity rating regularly triple that of Congress. However, if the systems are irreducible, what does that imply for attempts to reform?

Recognizing—and Modeling—Rapid Transitions

The difference between seemingly small decisions and those which make significant changes can be seen by analogy. Consider a ball resting on the flat top of a hill. It can roll around on top of the hill, but at some point, it will descend following a single downward path (among many potential paths). As it falls, it will speed up at the steepest part of the slope. However, the direction that it takes down the hill was determined when it began rolling down the side, even though at that point it was moving rather slowly. These slow changes near the top are most important in determining the path that it takes; once rolling down the hill, the ball's direction is set. Lamberson and Page (2012) warn that focusing on the moments at which maximal change occurs misses the importance of earlier forces, even if they capture less dramatic moments. The magnitude of events can deceive us. Early moments may exhibit only marginal changes in the characteristics of the system, but be critical for the future shape of it.[3]

Summary

These questions are by no means exhaustive of the directions that a research program in institutional robustness might take. Theories of institutional stability are nonadaptive. Robustness, in contrast, considers how institutions might adapt—either in their form or, in the case of a complex institutional space, in the relationship between components—to influence behavior to meet a social goal. Robustness studies also include a role for intentional (re-)design, a feature of institutions that sets them apart from ecological parallels.

---

[3] Introduction to the science can be found in Gunderson and Holling (2002), Scheffer et al. (2009), Page (2010), and Lamberson and Page (2012).

By moving away from a focus on stability and toward robustness, scholars interested in institutional dynamics gain insight into the interdependence between system components and nonlinear effects, including rapid transitions. The theoretical toolkit includes a set of logics about how institutions might change endogenously, including path dependence, evolution, and spillovers. Robust institutions are often characterized by modularity, diversity, and redundancy, and as the science of robust institutional design develops, scholars will be able to define further the conditions when each of these characteristics might improve system robustness.

Acknowledgments

I am deeply grateful to Rajiv Sethi for his suggestions to improve this article.