Institutionalism as a Methodology

Daniel Diermeier
Northwestern University

Keith Krehbiel
Stanford University

Thursday, May 23, 2002

Diermeier is IBM Professor of Regulation and Competitive Practice at the Kellogg Graduate School of Management, Northwestern University. Krehbiel is Edward B. Rust Professor of Political Science, Graduate School of Business, Stanford University. Part of this material was prepared while Krehbiel was a Fellow at the Center for Advanced Study in the Behavioral Sciences. Financial support was provided by the National Science Foundation #SES-9022192.

Abstract

We provide a definition of institutionalism and a schematic account that distinguishes between institutional theories (in which institutions are exogenous) and theories of institutions (in which some, but necessarily not all, institutions are endogenous). Our primary argument is that institutionalism in the contemporary context is better characterized as a method than as a body of substantive work motivated by the so-called chaos problem. Secondary arguments include the following. (1) While it is important to differentiate sharply between institutions and behavior, institutionalism as a method is neither inimical to behavioralism nor devoid of behavior. (2) When making the challenging transition from developing institutional theories to developing theories of institutions, it is essential to hold behavioral axioms fixed and to choose a form of equilibrium that exists for the class of games studied. (3) For most research programs today, a form of Nash Equilibrium has the requisite properties. The core, and structure-induced equilibria (SIE) that rely on the core, often lack the requisite properties.

Institutionalism as a Methodology

Daniel Diermeier and Keith Krehbiel

The seeds of formal analytic political science were planted a half-century ago with the publication of five major works: Kenneth Arrow's (1951) *Social Choice and Individual Values*, Duncan Black's (1958) *The Theory of Committees and Elections*, Anthony Downs's (1957) *An Economic Theory of Democracy*, William Riker's (1962) *Theory of Political Coalitions*, and Buchanan and Tullock's (1962) *Calculus of Consent*. Approximately midway between this germination of rational choice politics and the present, Kenneth Shepsle (1979) published "Institutional Arrangements and Equilibrium in Multidimensional Voting Models." Like the preceding foundational works, Shepsle's, too, proved to be seminal. It also marked a transition, inasmuch as it coincided with major turning points in two or more fields of political science. At one level, formal theorists as of 1980 began to worry less about analytic properties of relatively unconstrained social choice environments and began to think more about factors that seem to constrain decision-makers who engage in collective choice. At another level (yet, at the same time and, often, same place), legislative scholars were awakening to the prospect that, with judicious and empirically informed tailoring, formal models in the rational choice tradition could be used to gain a deeper understanding of legislative processes, behavior, and outcomes. The convergence of these two trends led to the birth of the *New Institutionalism*. Its mantra was that "institutions matter," and its mission evolved into reshaping not only legislative studies but also political science at-large.

The purpose of this essay is to clarify the nature and purpose of this research program as it has evolved during the last two decades. Our core argument is that *institutionalism*—when properly construed in today's research context—is more of a method than a mission. By this we mean that the aim of institutionalism circa 2000 is not to make a point that institutions matter because they somehow induce stability in otherwise chaotic collective choice situations. Rather, the aim of contemporary institutionalism is to guide inquiry into which of many more-or-less stable features of

collective choice settings are essential to understanding collective choice behavior and outcomes. Institutionalism, therefore, is particularly well-suited for comparative research, whether the institutional comparisons are cross-sectional or inter-temporal, or whether they are between committees or constitutions.

Our advocacy of institutionalism as a methodology is not intended merely to be food for thought in the ongoing history of political science. Rather, to the extent that institutionalism is compelling as a methodology, this perspective has important consequences for the praxis and progress of not only the institutionalist research program but also political science more broadly. This rather sweeping claim has both theoretical and empirical components.

At the theoretical level, the perspective has consequences for modeling strategies, the most noteworthy of which is the equilibrium concept used. A byproduct will be that some of the puzzles that have motivated rational choice research over the last few decades become, if not less puzzling, then at least less important. This is particularly true of the so-called chaos results in social choice theory. Specifically, once institutionalism is understood as a methodology, certain critiques of the institutionalist program cease to be valid. An example is the famous inheritability argument due to Riker (1980). Widely cited as a forceful critique of those early versions of institutionalism that interpret institutions as solutions to the instability problems of collective choice, the argument loses its force when applied to institutionalist studies based on alternative solution concepts.

At the empirical level, the practical consequence of institutionalism as a method is less subtle. Quite simply, empirical testing, whether in field studies or the laboratory, becomes an essential part of the institutionalist program. Because a central and immediate implication of the perspective of institutionalism as a method is that there is no such thing as "*the* theory of rational choice" (see, for example Green and Shapiro 1994), designing and conducting tests that discriminate between groups of theories that share methodological but not substantive assumptions is crucial to the viability of the research program.

To clarify the strengths of the methodological perspective on institutionalism, it is necessary to address several controversial questions, What is an institution? What do political scientists mean when they assert that "institutions matter?" What is institutionalism, and how does it relate to competing research programs such as behavioralism? And, most importantly, by what process can researchers develop more satisfactory theories of institutions?

We address these and related issues in five parts. Section 1 defines *institutions* and sets up the underlying thesis of the essay: that institutionalism is more constructively viewed as a method for research in collective choice processes than as a successor to—or competitor of—a prior research tradition such as behavioralism. Section 2 describes *institutional theories* as essential building blocks for more general theories. The method in this kind of research consists of analysis in which institutional features are taken as exogenous and behavioral postulates are fixed, and then compares equilibria that are generated under different institutional arrangements. Section 3 then takes the conceptually easy but operationally difficult step from institutional theories to *theories of institutions*, with the distinguishing characteristic that some institutional features become objects of collective choice. Section 4 discusses an important concern in the process of forming theories of institutions: the choice of equilibrium concept. Much of the discussion in this section concerns the comparative advantages of noncooperative-game-theoretic approaches versus core or structure-induced-equilibrium approaches. Section 5 revisits a well-known critique of institutionalism, Riker's Inheritability Argument, in light of the arguments developed in sections 1-3. Section 6 concludes.

1. Institutions

A political institution is a set of contextual features in a collective choice setting that defines constraints on, and opportunities for, individual behavior in the setting. In the context of legislative models, for example, such features include, but are not restricted to, the following. Who may and may not initiate proposals? In what order are proposals considered? Under what conditions can proposals be amended? Who has veto rights? Can vetoes be overridden? By what fraction of votes? By stipulating that

contextual features proscribe as well as prescribe individual behavior during processes of collective choice, this definition clearly allows for the possibility that "institutions matter." However, analysis of institutions does not (or should not) presuppose that different contextual features have different consequences for outcomes.

The crucial link between *institutions* (as contextual constraints) and *outcomes* (as consequences of collective choice) is *behavior*. While the line between institutions and behavior is not always easy to draw, it is well worth the effort to draw this line as sharply as possible to preserve the methodological distinction between the institution and the behavior that transpires within it. A rule of thumb, therefore, is to regard as an institution only contextual features that, in a given decision situation, are believed to constrain individual choices. Having done that and only that, notice that open but well-defined questions remain. Generally, the questions take the form: What are the consequences, if any, of the individual constraints on individual behavior and, in turn, on collective choices?

This proposed rule of thumb should not be construed as advocacy that the term institution should refer only to rigid, well-defined, constraining, immutable, formal, or structural features of collective choice.[1] Rather, we suggest only that the line be drawn comfortably on the firm side of mere patterns of behavior. If it is not, institutions and behavior become conceptually and analytically muddled, thereby making it exceedingly difficult to sort through what is assumed and what is derived in the ensuing formal argument.

This leeway in drawing the line between institutions and behavioral regularities becomes troubling only if one insists on an *ontological* distinction between institutions and behavior. As we argue in this essay, the distinction is better understood as a *methodological* one. For instance, depending on the research perspective, a congressional committee's gatekeeping authority may be interpreted as a constraint (e.g. if we want to study the likelihood that a certain bill will be passed) or as a behavioral

[1] Such features are sufficient by our definition but not necessary.

regularity (e.g. if we want to understand how legislative majorities decide on the internal organization of legislatures).[2]

Drawing the line between institutions and behavior seems easier in the study of elections than in the study of legislatures. Examples of the relevant institutions include the ballot structure, the rules for translating votes into seats, and district size etc.[3] In a given campaign, these rules can defensibly be assumed to be exogenous. This, in turn, allows the researcher to focus on the behavior of voters and candidates.[4] The distinction is less clear in the context of legislative models, however. Should rights of recognition or of bill introduction be considered an institution? And what about seniority norms? Does it matter whether a norm has never been violated?

We illustrate the advantages and disadvantages of the hard-line definition of institutions below. Meanwhile, to reiterate, we shall consider as institutions those contextual features of the decision-making setting that the researcher regards as essential to understanding how political actors behave in pursuit of their goals. Defined as such, institutions have the distinguishing feature of characterizing *incentives* for certain types of behavior as well as imposing *constraints* on such behavior. It cannot be stressed enough that, in this sense, behavior within the institution—not just the institution in isolation—determines whether institutions are outcome-consequential, or, as is more often uttered, whether institutions matter.

2. Institutional Theories

Generically, the question posed by institutional analysis is: How is the behavior of political actors and their collective choices influenced by incentives and constraints?

---

[2] In the language of modern philosophy of science our approach to institutions is "instrumentalist", not "realist".

[3] For a summary of the rich variety in electoral institutions see Cox (1997).

[4] On closer inspection similar issues emerge in the context of electoral models. After all, many countries can change their electoral institutions by passing a new law or amending the constitution. Recent

Following from the definition of an institution, an *institutional theory*—that is, a theory that seeks an understanding of the relationships between institutions, behavior, and outcomes—can be summarized as a four-step method depicted in Figure 1.

1. State and hold fixed *behavioral postulates* for political actors within the collective choice setting to be studied.
2. Characterize formally the *institutions* in effect (as defined in section 1).
3. Deduce the *behavior* that arises within the institutional setting, given the behavioral postulates, and characterize the *outcomes* that result from the behavior.
4. Evaluate the derived *empirical implications* using data.

Steps 1 and 2 of the method are the axioms and assumptions in the analysis, while steps 3 and 4 are the derivations and implications. Although the distinctions between axioms and assumptions, and between derivations and implications, are somewhat blurred, the more important distinction is between the pairs of terms: 1 and 2 versus 3 and 4. Steps 1 and 2 are *exogenous* within the context of a well-specified institutional theory. Steps 3 and 4, in contrast, are *endogenous*. Some stage-specific comments further clarify the concept of an institutional theory.

In stage 1, behavioral postulates may come in several forms, including myopic individual choice (e.g., sincere voting), bounded rationality (e.g. models of aspiration or satisficing), and rational choice (e.g., sophisticated voting, Nash equilibrium). Although institutional analysis can be conducted using any such postulate (or more specific postulates within these families), if the focus is on how institutions affect collective choices, it is crucial that behavioral postulates remain fixed and consistent within and across studies. (The reason for this becomes clearer below when we elaborate on step 4.) Because more often than not, institutionalists consider themselves to be rational choice theorists, we too will adopt that focus.

Rational choice theories postulate that decision-makers have well-defined preference orderings and that individual choices are consistent with this ordering. As of this writing, the most widely used equilibrium concepts are the core and Nash

examples include France, New Zealand, Italy, Japan, and Israel.

equilibrium (possibly with an appropriate refinement). Because a key element of institutional analysis is to vary institutional features while keeping the behavioral postulates constant, it is of great importance that the equilibrium concept be applicable to many collective choice settings. The tendency among rational choice theorists is not to be overly concerned with the literal descriptive accuracy of these behavioral postulates (e.g. Friedman 1953). In the context of institutional analysis, there is a practical methodological reason for keeping behavioral postulates simple. The focal object of study is the institution. Institutions are often complex. Keeping the mathematical model of behavior simple allows for a more transparent and tractable focus on the institutions that enter in stage 2.

In stage 2, researchers specify what is thought to be the essential contextual features of the arena to be studied. Perhaps even more so than stage 1, stage 2 modeling choices involve guesswork, particularly during the initial iterations of institutional theories in a given setting.[5] While the choice of features is, in practice, clouded by uncertainties about whether the model will indeed provide the desired account, the standard for a successful model is that it *isolates correctly* the *specific* institutional features that are determinants of behavior and outcomes. According to this standard, simplicity is a far greater virtue than the lack of descriptive accuracy is a vice.

Stage 3 often involves the heavy lifting in institutional modeling. Here equilibria are characterized and the mathematical implications of the model are derived. The model's implications may pertain to outcomes (e.g., whether a certain bill will pass) or to behavioral regularities (e.g., which legislator would vote for or against the bill). Different behavioral concepts can have variable degrees of power at this stage. Noncooperative models, for example, with Nash equilibrium as the relevant solution concept, generate both behavioral and outcome-related predictions, while cooperative solution concepts such as the core predict only outcomes.

[5] Shepsle and Weingast (1995), for example, refer to "first generation" models of legislatures (a.k.a. distributive theories) and defend them as limited in their predictions but necessary, or at least helpful, for subsequent generations of models.

Stage 4 consists of empirical assessments of the predictions of the model. Although this essay does not deal extensively with the issue of testing, its importance in rounding out the portrayal of the method of institutionalism is difficult to overstate. According to the methodological approach there is no intrinsic (but only a methodological) difference between institutions and behavioral regularities. Hence, any modeling decision (e.g., deciding which aspects of politics are to be modeled as exogenous and which are endogenous) is evaluated by whether it is useful to explain political phenomena; and this evaluation can be done only in empirical analysis. Likewise, the testing of institutional theories is essential in making confident judgments about whether the institutional theories are good building blocks for more general theories of institutions.

Finally, a comment about the iterative nature of development of institutional theories. Rational choice theorists are sometimes criticized for being unwilling to change their behavioral postulates (stage 1) as opposed to modifying other aspects of their model which, here, fall comfortably in the category of institutional assumptions (e.g., Green and Shapiro 1994). In the context of institutional analysis, this kind of criticism is misguided. The very aim of institutionalist analysis is to investigate different institutional settings. To do this, it is a methodological necessity to hold fixed the behavioral postulate. For example, if an institutional theory were to yield a falsified prediction and the researcher were to blame the behavioral postulate rather than the institutional assumptions for the falsification, then the next step would be to keep the same game form and alter the behavioral postulate. This may seem reasonable in isolation, but it is quite unreasonable to the degree that knowledge has accumulated elsewhere under the rubric of a single behavioral postulate, such as rational choice. If, in the presence of such knowledge, the behavioral postulate were abandoned, then the all prior institutional theories that contributed to the base of knowledge would have to be re-analyzed to gain comparability. In other words, making regular changes in behavioral postulates essentially guarantees that the field of study will fail to be cumulative. Likewise, as will be clearer below, changes in behavioral postulates would constitute a

severe setback for those who wish to make the transition from institutional theories to theories of institutions.

Of course, this argument is not meant to assert that the behavioral postulates given by rational choice theory are "the right ones." In the presence of persistent anomalies, even core assumptions of a research program should be reevaluated. The point is that, from an institutionalist perspective, the reluctance to give up a behavioral postulate is a methodological plus, not a minus. The same argument would apply to any other behavioral postulate.

3. Theories of Institutions

Institutional theories often elicit a somewhat misguided criticism for assuming that institutional features cannot be altered by the actors. The criticism is not empirically misguided because, often, decision-makers can and do change the structural arrangements under which they operate. However, the criticism is theoretically misguided inasmuch as it loses sight of the limited aim of institutional theories: structural features must be exogenous when the aim is to learn how and why contextual features affect choice processes. If the researcher wants to identify the institutional factors that explain a particular pattern of behavior, the institutional features simply cannot be modeled simultaneously as causes and consequences of that behavior.

This distinction between institutions (as exogenous) and behavior (as endogenous) is clear as long as we stay in the realm of institutionalist theories, but it becomes more complicated as we make the transition from *institutional theories* to *theories of institutions*.[6] The focus in a theory of institutions is to explain why some institutional features come into existence and persist while others are either nonexistent or transient. The defining characteristic of a theory of institutions is that *some* of the essential contextual features that were assumed to be constraining in the foundational institutional

[6] See Riker (1980), Shepsle (1986), and Calvert (1995) for essays that discuss the endogeneity of institutions.

theory become objects of choice within a *somewhat* more general theory of institutions. This necessarily partial endogenization of institutional features is what distinguishes an institutional theory from a theory of institutions.[7]

It should be obvious that a theory of institutions cannot exist without institutional theories. More precisely, in order to know why a certain institution exists, it is essential to know, with reasonable confidence, not only the consequences of the focal institution but also the consequences of alternative institutional arrangements that could have instead been crafted. The corresponding research strategy for theories of institutions embeds the steps for institutional theories as follows and as depicted in Figure 2:

A. State and hold fixed *behavioral postulates* believed to govern the choice from a set of specified institutions.

B. Conduct and/or embed *institutional analysis* (i.e., steps 1-4 above) for each institution identified in A.

C. Characterize formally the *second-order institutions* that constrain the choice of institutions defined in A.

D. Deduce the equilibrium *behavior* and *outcomes* of the institutional choice process in steps A-C

E. Evaluate the *empirical implications* using data.

The astute reader will have noticed that steps A-E are conceptually very similar to the four-step method from figure 1. This is not accidental, since a theory of institutions is, in effect, a meta-application of the institutional approach to the choice of institutions.

Step B underscores the point that theories of institutions cannot succeed without well-formulated and well-verified institutional theories. If institutional theories are not well-formulated (i.e., their assumptions are unclear; their solutions are not rigorous),

[7] This distinction is roughly consistent with what Shepsle (1986) called institutional equilibria and equilibrium institutions.

embedding them into theories of institutions will cause problems. If they are not well-verified, then the predictions about institutional choice are not likely to be corroborated either.

The key to avoiding confusion in constructing theories of institutions is to distinguish between levels or orders of institutions. In a theory of institutions the first-order institutions are objects of choice, but the second order institutions (i.e. the institutions that serve as constraints in the choice of first-order institutions) are exogenous. Of course, this process can be iterated further. Third-order institutions govern the choice of second-order institutions etc. But at every level, when we study the choice of institution of level k, the institutions at level $k+1$ are assumed to be constraining, thus exogenous.

An example may help to clarify this Russian doll approach. Consider the case of proposal power. A legislature may have a rule that states that only the chairman can make an initial proposal (call this the *chairman rule*). If we want to study the choice of bills under this institutional arrangement, then the chairman's proposal power is suitably modeled as an institution, i.e., it is considered fixed and unalterable during the decision on a given bill. However, we may ask why a legislature would give the chairman this procedural authority. *In this context* the chairman rule is only one of many that a legislature can choose. The decision on rules, however, again will be governed by other (second-order) institutions.

Note that the so-called chairman rule can be viewed either as a constraint (in the first-order choice problem) or as a choice alternative (in the second-order choice problem). But there is nothing circular or inconsistent in this view. Just as any set can also be viewed as an element of a larger set, any institution can also be viewed as the outcome of a higher order choice process,[8] provided the different levels of the institutional hierarchy are clearly distinguished.

[8] For discussions on whether this iterative process must stop eventually, i.e. whether there is a "last institution" see Calvert (1995) and Diermeier (1997).

For a recent example of this research strategy, consider Diermeier and Myerson's (1999) study of the effects of constitutional features on the internal organization of legislators using a two-stage noncooperative game (Step A). As a first step they generalize a vote-buying model proposed by Groseclose and Snyder (1996) to a variety of collective decision rules other than majority rule (Step B). The results from the solution of the legislative subgame are then used to analyze the choice problem faced by a chamber's pivotal voter(s) at the organizational stage on which decision rule to use. Chambers are assumed to maximize the expected payments by external lobbyists, but can commit only to organizational procedures within a chamber. Across chambers, no commitment is possible. Hence, in a multi-cameral setting (or when there is a president with veto power) the organizational choices of a chamber are interpreted as a noncooperative game between chambers (Step C). The Nash equilibria to this cross-chamber game (Step D) depend on the institutional details of the overall legislative process. For example, in unicameral settings, delegation of decision making power to a single actor (such as the cabinet a prime minister) is optimal. Bicameral and presidential systems, on the other hand, encourage the creation of internal veto players (such as, a committee that can block bills, but cannot pass them). While, Diermeier and Myerson do not conduct a detailed empirical analysis, their results can account for broad institutional differences between the internal organization of Congress and other presidential systems and unicameral, European systems (Step E).

This summary of the method of theories of institutions suggests that the distinction between constraints and equilibria is one that is sharp and critical methodologically, but fundamentally ambiguous— almost artificial—empirically. While this may be unsettling to many researchers, it should not be particularly bothersome. The key is that analysis of the study of institutions is, or should be, driven by specific topics or research interests. Institutionalism as a method proceeds most fruitfully when one takes as exogenous those contextual features of collective choice that are in some senses off limits to decision-makers

during the stages of collective choice studied.[9] However, which aspects of the decision process are off-limits is fundamentally a judgment call by the researcher.

Our conceptualization of theories of institutions does not necessarily preclude the use of different behavioral postulates at different levels of the hierarchy but it tends to discourage this approach. For example, researchers may want to assume sincere voting at the level of mass elections (here the electoral rules are first order institutions that constrain voters' choices), but allow for strategic voting when, for example, a constitutional convention decides on which electoral rule to use (in which case the procedural rules of the convention are second order institutions). However, much of the theoretical coherence of the institutionalist approach stems from the use of a general solution concept, and level-dependent solution concepts run the risk of undermining coherence. In the next section we will argue that this approach has direct consequences for research practice, especially in the theory of legislatures. Specifically, compared to its main alternative, structure-induced equilibrium theory, noncooperative game theory is a superior methodology for the institutionalist research agenda.

4. The Choice of Equilibrium Concepts

In many applications, researchers will not be overly concerned with steps 1 and (A)—the choice of behavioral postulates and corresponding solution concepts. Indeed, as we argued above, a conservative attitude towards solution concepts is essential for the cumulative nature of any institutionalist research program. To date, there have been two predominant approaches to construct theories of institutions. They differ not in intentions but rather in form. Specifically, the first set of attempts to model institutional choice were based on the solution concept of a *core* as manifested in Shepsle's notion of *structure-induced equilibrium*. The second, more recent set of attempts to model institutional choice as a *noncooperative game* uses various forms of *Nash Equilibrium* as the solution concept.

---

[9] For example, in their seminal work on agenda setting models, Romer and Rosenthal (1978) studied *local* education budget choice; not state constitutional reform.

Choices such as these may seem like mere matters of analytic convenience. The discussion that follows suggests this is not the case.

A. Structure-Induced Equilibrium

The historical success of the structure-induced-equilibrium (SIE) approach is largely due to its blend of analytical tools of social choice theory with a focus on institutional detail. These intellectual traditions resonated well both with legislative scholars and a new generation of formal theorists. The most comprehensive if not the most successful applications of the SIE approach are in the study of U.S. Congress. SIE models were among the first formalizations of phenomena—such as the committee system and jurisdictions—that already had been identified as important features of congressional decision making. Early applications also drew important conclusions about policy outcomes, e.g., that preference-outlying committees may lead to inefficient policies that favor special interests.[10]

In most instances, SIE models are presented as solutions to a collective choice problem suggested by McKelvey's intransitivity theorems (1976, 1979).[11] For example, as late as 1989, a leading new institutionalist referred ominously to living under "spell of McKelvey and his colleagues" when providing a motivation for an institutionalist theory (Weingast 1989, 795). Dubbed the chaos theorem, McKelvey's result is that one can construct a sequence of alternatives between any two points in a multidimensional space that has the property that myopic majority-rule voting will lead from the starting to the ending point in the sequence. The result has many interpretations, but the standout favorite of SIE researchers is that models of McKelvey and Schofield's types predict "chaotic," discontinuous changes in policy outcomes (see, for example, Shepsle 1986).[12]

---

[10] See Weingast and Marshall 1988, Shepsle and Weingast 1984, and Weingast, Shepsle and Johnsen 1981.

[11] This notion of "solution to the instability problem" clearly has a functionalist connotation, insofar as it suggests that the institutions emerge or are invented without specifying the individual actors' preferences in the case of institutional choice.

[12] Diermeier (1999) criticizes such an interpretation of McKelvey's result. The argument is that McKelvey's Theorem should not be interpreted as a theory that tries to explain or predict observable

In contrast to social choice theory, Shepsle's institutional theory embodies structural features that impose sufficient order on collective choice that stable outcomes result.[13] This solution—understandably called a *structure-induced equilibrium*—epitomizes what researchers in this tradition soon proclaimed as the New Institutionalism.

SIE theories were not presented as methodological innovations over social choice theory—indeed, Shepsle's seminal article relies heavily on social choice concepts and notation. Nor were SIE works merely theoretical critiques of prior models. They were also—and perhaps most directly—challenges to social choice theory on *empirical* grounds. That is, SIE theorists asserted that McKelvey's results are inconsistent with observed regularities. Social choice theory was interpreted as predicting the absence of stability, while, in contrast, the empirical regularities observed by congressional scholars was that outcomes were predictable and stable (Shepsle 1986). Having concluded on empirical grounds that McKelvey's theory needed modification, Shepsle (1979) chose to introduce analytic forms of institutions that resembled apparent real-world constraints on behavior. A key example was the notion of a jurisdictional system that, in his formulation, partitions a multidimensional choice space into subsets of single issues that are voted on one at a time. As Kramer (1972) had demonstrated earlier, a core exists in issue-by-issue voting. The key methodological idea of SIE theory thus consists of transforming a social choice problem in which the core does not exist (multi-dimensional choice spaces) into a more structured problem in which the core does exist. Note, however, that such a strategy maintains the core as the equilibrium concept.

In core-based institutionalism, institutions are structural features that are abstract in formal models but real and constraining features in actual collective choice settings. To researchers in this field, institutions matter in the sense that they are necessary to make predictions in settings in which otherwise no prediction is possible. While the SIE

---

behavior, but as a conceptual investigation of the apparatus used to construct such explanatory theories. Like the core nonexistence results (Plott 1967, McKelvey and Schofield 1987), the McKelvey theorems define the domain where certain explanatory concepts can be used, but make no claims about a specific application of these concepts to a particular piece of political reality.

[13] See Krehbiel 1987 for an extensive but nontechnical review of Shepsle's model.

literature has clear conceptions of equilibrium (the issue-by-issue core) and institutions (assignments of issue dimensions to groups of individuals), it does not possess an explicit behavioral model. Issue-by-issue cores depend only on (collective) preference relations. While this may be an advantage at the theoretical level, it makes empirical evaluations harder, since detailed information about the outcome of political processes must be obtained to judge the empirical adequacy of a SIE model.

Another problem with core-based institutionalism pertains to the domain of SIE in institutional analysis. In contrast to Nash equilibrium, issue-by-issue cores exist only in very restricted circumstances. This limits the applicability of SIE theory beyond the narrow realm of issue-by-issue voting in legislatures. Technically, this domain is dictated by the applicability of a version of Kramer's (1972) fixed point argument. In the paradigmatic case of legislative committees, for instance, the SIE methodology cannot effectively accommodate matters of jurisdictional reform, committee assignments, and choice of rules. The problem is a manifestation of Riker's (1980) famous inheritability problem. Even if SIE exist for all of the institutional arrangements in, say, figure 2, and if the set of arrangements from which actors choose is moderately diverse, the induced social preference ordering over the SIE will yield an empty core. If not, the absence of a core in the broader theory of institutions means, in effect, that there is no such theory—at least not one capable of deriving predictions. In short, the chaos problem is inherited by the choosers of institutions in the larger core-based game.

Inheritability leads to an immediate problem for SIE theory given its reliance on the core as its solution concept. It is well-known that in spatial models where the outcome space has at least two dimensions the core is almost always empty (Plott 1967, Schofield 1983, McKelvey and Schofield 1987). Core non-existence results, however, are not restricted to spatial models, but also occur in the case of more than two alternatives in the finite case (Nakamura 1979). This makes it highly doubtful that SIE theory can *in principle* come up with a model that would yield ex-post-veto arrangements or other structural features as equilibrium institutions. Indeed, in the cases where such an equilibrium analysis has been attempted, it was negative; that is, in general, it yielded an empty core (Austen-Smith and Banks 1990, Laver and Shepsle 1990, Diermeier 1997).

Without an explicit model that has constraining structures as equilibrium outcomes, however, the verbal defense of structural features lacks a rigorous foundation.

In conclusion, SIE-based methods are not well-suited for theories of institutions, because the second-order choice environments are unlikely to satisfy the stringent equilibrium existence conditions identified by abstract social choice theory. In cases where equilibria do not exist, the theory makes no predictions. Second, following Riker's inheritability argument, because there can be no successful SIE-based theory of institutions, the rationale for treating institutions, such as committees or jurisdictions, as binding (in the sense of the inducing a core) is undermined. The common cause to both of these problems is the generic non-existence of the key solution concept underlying SIE, the core. In summary, these facts suggest that it would be wise to abandon social-choice based concepts in favor of equilibrium concepts that generate predictions in many circumstances.

B. Noncooperative Game Theory

Since the mid-eighties noncooperative game theory has emerged as one of the dominant formal approaches in theories of institutions. The common denominator of this approach is that it views political interaction as a noncooperative game. This allows a clear formal representation of both institutional analysis and the theory of institutional choice. Let us consider institutional analysis first.

As its behavioral postulate, noncooperative game theory assumes that all actors choose best-response strategies. That is, they choose the strategy that (given the other actors' strategies) leads to an outcome that is maximal in their preference ordering. Institutions are modeled as game-forms; a complete description of all available strategies for all players is provided; and a function identifies the outcomes of all possible strategy combinations. For the purpose of the analysis, the game form is considered fixed and exogenous. As already pointed out, this is a methodological asset: what is assumed to be exogenous in institutional analysis at first can be the object of institutional choice subsequently. Finally, the equilibrium concept used is Nash-equilibrium (possibly with a refinement), which captures the intuition that predicted outcomes must be mutual best

responses. That is, in equilibrium no actor has an incentive to choose any other than her prescribed strategy.

Noncooperative game theory has two main advantages for the study of institutions. First, Nash equilibria exist for almost all game forms. This implies that game theory can be used to model many different institutional settings. Second, the qualitative features of Nash equilibria are highly sensitive to the precise details of the game forms. From the very beginning, this fact precludes any notion of an "institution-free" theory. Rather, the potential fruitfulness of the game-theoretic approach stems from the simple fact that it explicitly models some features of political institutions, and, therefore, highlights *how* and *why* institutions matter.[14]

As an example, consider the use of the Romer-Rosenthal agenda setter model (Romer and Rosenthal 1978, 1979). While the model has been widely used in legislative studies, it originally was devised to study school bond referenda. The structure of the basic model is straightforward. It features an agenda setter, who can make a take it or leave it offer, and a pivotal actor. If the pivotal actor rejects the proposal, an exogenously given status quo point is implemented. Otherwise, the setter's accepted proposal is the outcome.

Historically, the importance of the Romer-Rosenthal model lies in its connection to Black's median voter theorem. In a unidimensional model (with symmetric, single-peaked preferences and an odd number of voters) the median voter's ideal point is the core. The median voter theorem thus predicts the median voter's ideal policy as the outcome of the collective choice process. On the other hand, the solution of the Romer-Rosenthal model (that is, its unique subgame perfect equilibrium outcome), on the other hand, predicts that the chosen policy will be biased away from the median voter's ideal point towards the ideal point of the proposer, unless proposer and median have the same preferences. Thus, the Romer-Rosenthal is another key example of why institutions matter in the case of a closed rule. An institutionalist approach to the Romer-Rosenthal model, however, would not stop here, but rather would compare different institutions in terms of their

[14]One may argue that the explanatory power of game-theory is also limited by the very same features. Many games have multiple equilibria and sometimes the analysis seems to depend too much on the

consequences. Baron (1996), for example, constructed a noncooperative multi-period model with random proposer selection to capture an open rule proposal process. In his model, policies converge to the median voter over time. The point of this model is that now both the closed proposal rule studied by Romer and Rosenthal and an open process can be captured by the same methodology, noncooperative game theory.

There is no need here to repeat the spirited debate on whether models with or without such proposal power are appropriate representations of, say, congressional decision making.[15] It suffices to note that any model with agenda setting power needs eventually to answer the question of why a chamber majority would grant such procedural prerogatives to some of its members. The debate about the answer is, in effect, a debate about the game form that best represents congressional decision-making. Different game forms are then suggested, solved, and then compared with empirical regularities. The existence of Nash equilibria in almost all applications therefore has important consequences for the study of institutions. Specifically, it allows researchers to vary the institutional details by varying the game form and then comparing the resulting equilibria.[16]

At the next level, game theory can be used to analyze institutional choice. Explicit formal examples of equilibrium institutions are still rare. Calvert (1995) is among the most explicit adopters and vigorous defenders of the equilibrium-institutions program. Other works embody some of its features, but they do not explicitly define institutions as equilibria. Diermeier (1995), for example, presents an explicit theory that states conditions under which deference to committees emerges endogenously. Thus, if one regards deference as an

---

details of the game-form, especially in bargaining models.

[15] Summaries can be found in Krehbiel (1988) and (1991).

[16] As an example, note that the use of the Romer-Rosenthal model (or any other game theoretic model) is not restricted to unidimensional choice space. Like any other game that explicitly models agenda setting works also in multi-dimensional settings. True, the debate on proposal power has been mainly restricted to uni-dimensional models, since it allows a direct comparison between the model's prediction and the median voter theorem. (It is also technically much easier.) However, the model can also be used in multi-dimensional choice spaces as long as preferences are continuous and convex. In this case, any given reversion policy induces a star-shaped, compact set of outcomes that some decisive coalition prefers over the reversion policy. A proposer can then simply suggest his most favored alternative from this set and be sure that such a proposal will always be accepted.

institution, this is an example of an explicit theory of an equilibrium institution. Shepsle and Nalebuff (1990) as well as McKelvey and Riezman (1992) analyze seniority norms in a related fashion. Similarly, Gilligan and Krehbiel (1987) model conditions under which a closed rule is adopted by a collective-choice body and constitutes a "procedural equilibrium," while Diermeier and Myerson (1999) construct a bargaining model between chambers to explain cross-national differences in the internal organization of legislatures.

In summary, two properties of noncooperative game theory facilitate both institutionalist theories and theories of institutions. First, equilibria exist under very general conditions. Second, games in extensive form can be decomposed into subgames, where each subgame may correspond to a different institution. For example, in the Gilligan-Krehbiel model, congressional decision-making under closed and open rules corresponds to two subgames under incomplete information, each with distinct informational and distributive properties. The choice of the rule by the chamber can be modeled as a choice between two subgames determined by the floor median's expected equilibrium payoff in each subgame. In this manner, the institutional hierarchy discussed in steps (A) to (E) is formally represented as nested subgames of increasing complexity.

5. Riker Revisited

Since its inception, the new institutionalism has been the subject of criticism, both within and outside of rational choice theory. As noted above, perhaps the most famous rational-choice-based critique is due to Riker (1980), who questions the theoretical foundations of institutionalism. At the level of institutional choice, argues Riker, SIE theories face the same issue of instability as social choice theory. That is, once we apply the same reasoning to choosing institutions, the problem of non-existence of equilibia

reemerges. Riker summarizes this inheritability problem as follows:

> In that sense rules or institutions are just mere alternatives in the policy space, and the status quo of one set of rules can be supplanted with another set of rules. ... If institutions are congealed tastes, and if tastes lack equilibrium, then so also do institutions, except for short-run events. (Riker 1980, 445)

From this argument, Riker concludes that, since an "equilibrium of tastes" is absent in political domains, politics (in contrast to economics) may not be predictable at all.

> In what seems to me a deeper sense, however, politics is *the* dismal science because we have learned from it that there are not fundamental equilibria to predict. In the absence of such equilibria we cannot know much about the future at all... (Riker 1980, 443)

Although the inheritability problem was a provocative-to-devastating blow to SIE-based institutionalism (Diermeier 1997), it does not follow from Riker's critique that there is a similar problem with the institutionalism research program in general. The weak link in this chain of logic is Riker's tacit assumption that the political substance of institutions was inextricably linked to the political-science methodology of institutionalism in the structure-induced-equilibrium tradition. Like Shepsle, Riker presupposes that the core is the appropriate explanatory concept for a theory of politics. Thereafter, their more explicit disagreement is on whether SIE models do or do not constitute a viable solution to the chaos problem.

Riker and Shepsle's point of disagreement has received much more attention than their point of agreement, but each misses the main point. The main point is that no longer is there a need to rely on the core as the central solution concept in the analysis of institutions. In other words, the common presupposition of these early, formal institutionalists is not defensible. The more promising alternative to core-based analysis is noncooperative game theory,

because its solution concepts are not afflicted with problems of nonexistence, and its flexibility avails the method to theory development in a remarkably wide range of social settings.

To recapitulate, an institutional theory of politics grounded in noncooperative game theory abandons Riker's notions of an "equilibrium of tastes" and "institution-free" politics. It does not follow, however, that the search for a general theory of politics must be abandoned as well, nor that political behavior is inherently unpredictable. Rather, the alternative methodology of institutionalism can and should be used in the pursuit of increasingly general theories of politics, albeit theories that are institutional (in the sense of game form) at every level of analysis. Therefore, to Riker's suggestion that the "dismal" problems political scientists confront are due to deep truths about politics, we would respond: the problem is not politics, itself, but rather the insistence by political scientists on an equilibrium concept that fails to generate predictions in most political domains. After a half-century of mixed success in pursuing the promises of formal theories of politics, there is now a demonstrably newer and better game in town.

5. Conclusion

This paper has provided an account of institutionalism as a methodology. A central argument is that there is no intrinsic difference between robust behavioral regularities and institutions, since the adoption and maintenance of institutions itself are functions of behavior in collective choice processes. As such, institutions are best interpreted as theoretical constructs. Which features of reality are to be treated as institutions is ultimately a modeling decision that should be motivated by a theory's explanatory power. This approach allows researchers not only to compare political institutions, but also to construct theories of institutions. That is, to investigate why certain institutions exist in the first place. General theories of institutions, then, correspond to a hierarchy of institutional models,

where lower level institutions are chosen according to the constraints imposed by higher level institutions.

The methodological perspective of institutionalism has direct consequences for research practice. First, since modeling choices are driven by explanatory power, empirical testing becomes an integral part of the institutionalist research program. Second, the need for an equilibrium concept that is general and flexible enough to model a wide variety of institutions and institutional choice strongly favors noncooperative game theory over structure-induced equilibrium theory. Third, Riker's famous inheritability argument does not present a problem for institutionalism, but it does undermine the use of structure-induced equilibrium as a promising institutionalist methodology.

References

Arrow, Kenneth J. 1951. Social Choice and Individual Values. New York: Wiley.

Austen-Smith, David, and Jeffrey S. Banks. 1990. "Stable Governments and the Allocation of Policy Portfolios." American Political Science Review 84:891-906.

Jonathan Bendor, and Terry M. Moe. 1986. Agenda Control, Committee Capture and the Dynamics of Institutional Politics. The American Political Science Review. 80(4): 1187-1207.

Baron, David. 1996. A Dynamic Theory of Collective Goods Programs. The American Political Science Review 90(2):316-30.

Black, Duncan. 1958. The Theory of Committees and Elections. London: Cambridge University Press.

Buchanan, James M., and Gordon Tullock. 1962. The Calculus of Consent. Ann Arbor: University of Michigan Press.

Calvert, Randall. 1995. "The Rational Choice Theory of Social Institutions: Cooperation, Coordination, and Communication." In Jeffrey Banks and Eric Hanushek, eds. Modern Political Economy. Old Topics, New Directions. 216-28. New York: Cambridge University Press.

Gary W. Cox. 1997. Making Votes Count : Strategic Coordination in the World's Electoral System. New York: Cambridge University Press.

Diermeier, Daniel. 1995. Commitment, Deference, and Legislative Institutions. American Political Science Review 89: 344-55.

Diermeier, Daniel. 1997. Explanatory Concepts in Positive Political Theory. Mimeo. Graduate School of Business. Stanford University.

Diermeier, Daniel, and Roger B. Myerson. 1999. "Bicameralism and Its Consequences for the Internal Organization of Legislatures," American Economic Review. 1999. 89(5):1182-1196.

Downs, Anthony. 1957. An Economic Theory of Democracy. New York: Harper and Row.

Friedman, Milton. 1953. Essays in Positive Economics. Chicago: University of Chicago Press.

Green, Donald, and Ian Shapiro. 1994. Pathologies of Rational Choice Theory: A Critique of Applications in Political Science. New Haven: Yale University Press.

Groseclose, Timothy, and James M. Snyder, Jr. 1996. Buying Supermajorities. American Political Science Review. 90(2): 303-315.

Kramer, Gerald H. 1972. Sophisticated Voting Over Multidimensional Choice Spaces. Journal of Mathematical Sociology 2: 165-180.

Krehbiel, Keith. 1987. Sophisticated Committees and Structure-Induced Equilibria in Congress. In Congress: Structure and Policy, ed. Mathew McCubbins and Terry Sullivan. New York: Cambridge University Press.

Krehbiel, Keith. 1987. Why Are Congressional Committees Powerful? American Political Science Review 81: 929-35.

Krehbiel, Keith. 1988. Spatial Models of Legislative Choice. Legislative Studies Quarterly 8: 259-319.

Krehbiel, Keith. 1991. Information and Legislative Organization. Ann Arbor: University of Michigan Press.

Laver, Michael and Kenneth W. Shepsle. 1990. Coalitions and Cabinet Government. American Political Science Review 84: 873-890.

McKelvey, Richard D. 1976. Intransitivities in Multidimensional Voting Models and Some Implications for Agenda Control. Journal of Economic Theory 12:472-482.

McKelvey, Richard D. 1979. General Conditions for Global Intransitivities in Formal Voting Models. Econometrica 47: 1086-1112.

Richard D. McKelvey and Raymond Riezman. 1992. Seniority in Legislatures. The American Political Science Review. 86(4): 951-965.

McKelvey, Richard D., and Norman Schofield. 1987. Generalized Symmetry Conditions at a Core Point. Econometrica 55: 923-34.

Nakamura, Kenji. 1979. The Vetoers in a Simple Game with Ordinal Preferences. International Journal of Game Theory 8: 55-61.

Plott, Charles R. 1967. A Notion of Equilibrium and its Possibility Under Majority Rule. American Economic Review 57: 787-806.

Riker, William H. 1962. The Theory of Political Coalitions. New Haven: Yale University Press.

Riker, William H. 1980. Implications from the Disequilibrium of Majority Rule for the Study of Institutions. American Political Science Review 74: 432-46.

Romer, Thomas, and Howard Rosenthal. 1978. Political Resource Allocation, Controlled Agendas, and the Status Quo. Public Choice 33: 27-43.

Romer, Thomas, and Howard Rosenthal. 1979. The Elusive Median Voter. Journal of Public Economics 12: 143-70.

Shepsle, Kenneth A. 1979. Institutional Arrangements and Equilibrium in Multidimensional Voting Models. American Journal of Political Science 23: 27-60.

Shepsle, Kenneth A. 1986. Institutional Equilibrium and Equilibrium Institutions. In Political Science: The Science of Politics, ed. Herbert Weisberg. New York: Agathon Press.

Shepsle, Kenneth A., and Barry Nalebuff. 1990. The Commitment to Seniority in Self-Governing Groups. Journal of Law, Economics, & Organization 6: 42-72.

Shepsle, Kenneth A., and Barry R. Weingast. 1984. Political Solutions to Market Problems. American Political Science Review 78: 417-34.

Weingast, Barry R. 1989. Floor Behavior in the U.S. Congress: Committee Power under the Open Rule. American Political Science Review 83: 795-816.

Weingast, Barry R., and William Marshall. 1988. The Industrial Organization of Congress. Journal of Political Economy 96: 132-63.

Weingast, Barry R., Kenneth A. Shepsle, and Christopher Johnsen. 1981. The Political Economy of Benefits and Costs: A Neoclassical Approach to Distributive Politics. Journal of Political Economy 89:642-64.

Figure 1. The method of building institutional theories

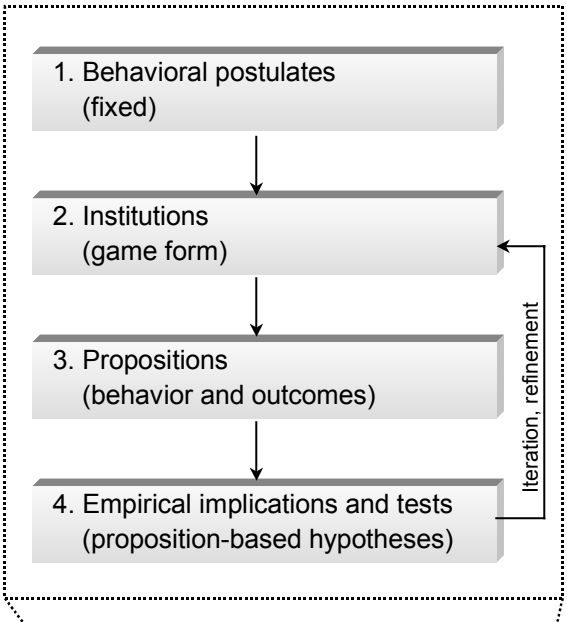

Figure 2. The method of building theories of institutions