Journal of Theoretical Politics

Institutionalism as a Methodology

Daniel Diermeier and Keith Krehbiel

The online version of this article can be found at:
http://jtp.sagepub.com/cgi/content/abstract/15/2/123

Published by:
$SAGE Publications

http://www.sagepublications.com

Additional services and information for *Journal of Theoretical Politics* can be found at:

Email Alerts: http://jtp.sagepub.com/cgi/alerts

Subscriptions: http://jtp.sagepub.com/subscriptions

Reprints: http://www.sagepub.com/journalsReprints.nav

Permissions: http://www.sagepub.com/journalsPermissions.nav

Citations (this article cites 10 articles hosted on the
SAGE Journals Online and HighWire Press platforms):
http://jtp.sagepub.com/cgi/content/abstract/15/2/123#BIBL

INSTITUTIONALISM AS A METHODOLOGY

Daniel Diermeier and Keith Krehbiel

ABSTRACT

We provide a definition of institutionalism and a schematic account that differ-
entiates between institutional theories (in which institutions are exogenous)
and theories of institutions, in which some (but not necessarily all) institutions
are endogenous. Our primary argument is that institutionalism in the contem-
porary context is better characterized as a method than as a body of substan-
tive 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 presupposes a well-defined beha-
vioral concept. (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 while the core, and structure-
induced equilibria (SIE) that rely on the core, often lack the requisite
properties.

KEY WORDS ● formal models ● institutions ● institutional ● methodology ●
rational choice

The seeds of formal analytic political science were planted a half-century ago
with the publication of five major works: Ken 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) *The Theory of Political Coalitions* and Buchanan and
Tullock's (1962) *The Calculus of Consent*. Approximately midway between
this germination of rational choice politics and the present, Kenneth Shepsle
(1979) published 'Institutional Arrangements and Equilibrium in Multi-
dimensional 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.[1] 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.

---

1. For other accounts of institutionalism see, for example, Hall and Taylor (1996).

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 method.

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 research 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*, the distinguishing characteristic of which is 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 non-cooperative-game-theoretic approaches versus core or structure-induced-equilibrium (SIE) 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 contains concluding comments.

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 con-
ditions can proposals be amended? Who has veto rights? Can vetos be over-
ridden? By what fraction of votes? By stipulating that contextual features
proscribe as well as prescribe individual behavior during processes of collec-
tive choice, this definition clearly allows for the possibility that 'institutions
matter'. However, analysis of institutions does not (and should not) pre-
suppose that different contextual features have different consequences for
outcomes.

The crucial link between *institutions* (as contextual constraints) and *out-
comes* (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, note that
open but well-defined questions remain. Generally, the questions take the
form: What are the consequences, if any, or the individual constraints on
individual behavior and, in turn, 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.[2] 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 con-
ceptually 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 regu-
larities 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* distinction. That is, depending on
the research perspective, for instance, a congressional committee's gate-
keeping 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 regularity
(e.g. if we want to understand how legislative majorities decide on the
internal organization of legislatures).[3]

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

---

2. Such features are sufficient by our definition but not necessary.

3. In the language of modern philosophy of science our approach to institutions is 'instrumen-
talist' not 'realist'.

seats and district size, etc.[4] 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.[5] 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 or other regularities such as apparent gate-keeping by standing committees? Should these phenomena be characterized as institutions, even if they might, and occasionally do, fail to constrain behavior?

We illustrate the advantages and disadvantages of the hard-line definition of institutions later. 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 are the behavior of political actors and their collective choices influenced by incentives and constraints? 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. Define 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 postulate and characterize the *outcome* that results from the behavior.
4. Compare the derived implications with empirical regularities and *data*.

---

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

5. On closer inspection similar issues emerge in the context of electoral models. After all, countries can change their electoral institutions by passing a new law or amending the constitution. Recent examples include France, New Zealand, Italy, Japan and Israel.

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 later 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 equilibrium (possibly with an appropriate refinement). Because a key element of institutional analysis is to vary institutional features while keeping the behavioral postulate constant, it is of great importance that the equilibrium concept is applicable in many collective choice problems. 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 detailed, yet tractable, focus on the institutions which enter in stage 2.

In stage 2, researchers specify what are 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.[6] 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

---

6. Shepsle and Weingast (1994), 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.

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 contains the actual work in institutional modeling. Here the mathematical implications of the model are derived, equilibria are characterized, etc. The model's implications may only 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 differ in this respect. Non-cooperative models, for example, with Nash equilibrium as the relevant solution concept generate both behavioral and outcome-related predictions, while cooperative solution concepts like the core only predict outcomes.

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. which aspects of the political phenomenon are to be modeled as exogenous and which as endogenous) is evaluated by whether it is useful to explain political phenomena; and this evaluation can only be done in empirical analysis. Likewise, the testing of institutional theories is essential in making confident judgements about whether the institutional theories are good building blocks for more general theories.

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 postulate (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 all of the prior institutional theories that contributed to the base of knowledge would have to be re-analyzed to gain comparability. In other words, regular changes in behavioral postulates essentially guarantees that the field of study

will fail to be cumulative. Likewise, as will be clearer later, 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 does not demonstrate 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 re-evaluated. The point is that from an institutionalist perspective, the reluctance to give up a behavioral postulate is a methodological virtue, not a vice. The same argument would apply to any other behavioral postulate.

3. Theories of Institutions

Institutional theories often elicit a somewhat misguided criticism that stems from the fact, for the purpose of the analysis, that the theories assume that institutional features cannot be altered by the actors. The criticism is not empirically misguided because decision-makers often 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 is fairly clear as long as we stay in the realm of institutionalist theories but becomes more complicated as we make the transition from *institutional theories* to *theories of institutions*.[7] The focus in a theory of institutions is to explain why some institutional features come into existence, and persist, while others are either non-existent 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 theory become objects of choice within a *somewhat* more general theory of institutions. The necessarily partial endogenization of institutional features is what distinguishes an institutional theory from a theory of institutions.[8]

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

---

7. See Riker (1980), Shepsle (1986) and Calvert (1995) for essays that discuss the endogeneity of institutions.
8. This distinction is roughly consistent with what Shepsle (1986) called institutional equilibria and equilibrium institutions.

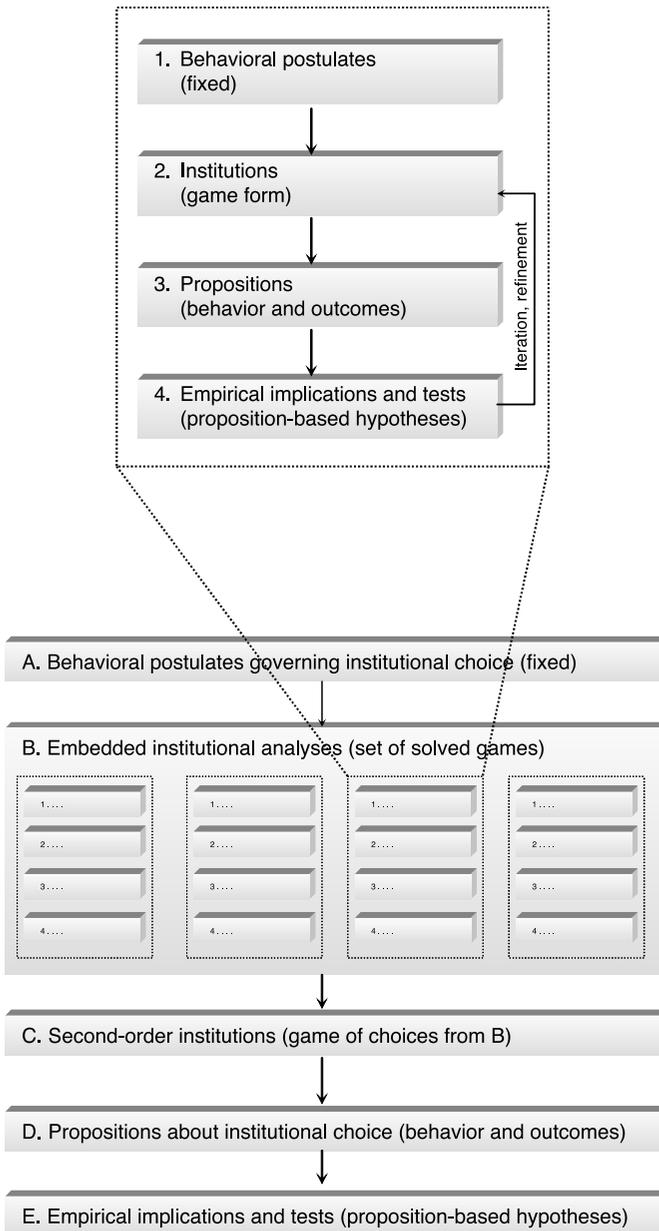Figure 1. The Method of Building Institutional Theories

Figure 2. The Method of Building Theories of Institutions

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:

1. Define and hold fixed behavioral postulates for political actors in collective choice process in which well-defined institutions are explicit objects of choice.
2. Conduct and/or embed institutional analysis (i.e. steps 1–4 as before) for each institution identified in A.
3. Characterize formally the (second-order) institutions that constrain the choice of institutions defined in A.
4. Deduce the set of institutions that will be chosen given steps A–C.
5. Compare the derived implications with empirical regularities and data.

The astute reader will have noticed that steps A–E are nothing but an extension of the four-step method from Figure 1. This is not an accident: we argue that an (institutionalist) theory of institutions is nothing else but an application of the institutional approach to the choice of institutions. Step B underscores the obvious 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), 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 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 constraints, 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 these powers. *In this context* the chairman rule is only one of many that a legislature can choose. The decision rules, however, will be governed by other (second-

order) institutions. For example, the eldest member of the chamber may be the sole proposer of first-order rules, and so forth.

Note that the chairman rule can either be viewed 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,[9] provided the different levels of the institutional hierarchy are clearly distinguished.

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 non-cooperative 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 only commit to organizational procedures within a chamber. Across chambers, no commitment is possible. Hence, in a multicameral setting (or when there is a president with veto power) the organizational choices of a chamber are interpreted as a non-cooperative 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 or prime minister) is optimal. Bicameral and presidential systems, in contrast, encourage the creation of internal veto-players (like a committee that can block bills but cannot pass them). While, Diermeier and Myerson (1999) 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 methodologically sharp and critical but empirically ambiguous if not artificial. While this may be unsettling to many researchers, it should not be particularly bothersome. The key is that analysis 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

---

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

choice that are, in some senses, off limits to decision-makers during the stages of collective choice studied.[10] However, which aspects of the decision process are off limits is fundamentally a choice made by the researcher.

Our presentation of theories of institutions does not preclude the use of different behavioral postulates at different levels of the hierarchy. 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, e.g., a constitutional convention decides on which electoral rule to use (so here the procedural rules of the conventions are second-order institutions). However, as we will argue later, much of the theoretical coherence of the institutionalist approach stems from the use of a general solution concept, which would be compromised by level-dependent solution concepts.

This concludes our overview of the institutionalism as a methodology. In the next section we will argue that this approach has direct consequences for research practice, especially in the theory of legislatures. Specifically, we argue that (compared to its main alternative, structure-induced equilibrium theory) non-cooperative 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 earlier, 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 rested 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 *non-cooperative game,* uses various forms of *Nash equilibrium* as the solution concept. Choices such as these may seem like mere matters of analytic convenience. However, the discussion that follows suggests this is not the case.

---

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

A. Structure-induced equilibrium

The historical success of the structure-induced-equilibrium (SIE) approach is largely due to the fact that it combined the analytical tools of social choice theory with a focus on institutional detail. This blend of intellectual traditions resonated both with legislative scholars and a new generation of formal theorists. The most comprehensive if not successful applications of the SIE approach are to the study of the US 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, such as that outlying committees may lead to inefficient policies that favor special interests.[11]

In most instances, SIE models are presented as solutions to a collective choice problem suggested by McKelvey's intransitivity theorems (1976, 1979).[12] For example, as late as 1989, a leading new institutionalist referred ominously to living under the 'spell of McKelvey and his colleagues' when providing a motivation for an institutionalist theory (Weingast, 1989: 795). Dubbed 'chaos theorems', 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).[13]

In contrast to social choice theory, Shepsle's theory embodies structural features that impose sufficient order on collective choice that stable outcomes result.[14] This solution – understandably called an *SIE* – epitomizes what researchers in this tradition soon proclaimed as the New Institutionalism.

---

11. See Weingast and Marshall (1988), Shepsle and Weingast (1984) and Weingast et al. (1981).

12. 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.

13. One of us has recently criticized such an interpretation of McKelvey's result (Diermeier, 1997). The argument there is that McKelvey's Theorem should not be interpreted as a theory that tries to explain or predict observable behavior but as a conceptual investigation of the apparatus used to construct such explanatory theories. Like the core non-existence 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.

14. See Krehbiel (1988) for an extensive but non-technical review of Shepsle's model.

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, congressional scholars observed 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, cores exist in issue-by-issue voting. The key methodological idea of SIE theory thus consists in transforming a social choice problem in which the core does not exist (multidimensional 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 was possible. Note, however, that while the SIE 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 an 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 reason is simply a manifestation of Riker's (1980) famous 'inheritability' problem. Even if SIE exists for all of the institutional arrangements in, say, Figure 2, if this 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 actors who choose institutions in the larger core-based game.

But this leads to an immediate problem for SIE theory. 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; 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). But 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). But without a model that has constraining structures as equilibrium outcomes, the verbal defense of structural features lacks a rigorous foundation.

This has two consequences. First, SIE-based approaches are not well suited for theories of institutions since the second-order choice environments are unlikely to satisfy the stringent equilibrium existence conditions identi-fied by abstract social choice theory. But in cases where equilibria do not exist, the theory makes no predictions. Second, following Riker's inherit-ability argument, because there can be no successful SIE-based theory of institutions, the rationale for treating jurisdictional assignments as binding (in the sense of the inducing a core) is undermined. The common cause of both of these problems is the generic non-existence of the key solution con-cept underlying SIE, the core. This suggests that it would be wise to abandon social-choice-based concepts in favor of equilibrium concepts that generate predictions in many circumstances. Here, the preferred alternative choice among rational choice theorists has been non-cooperative game theory with the Nash equilibrium as the relevant solution concept.

B. *Non-cooperative game theory*

Since the mid-1980s non-cooperative 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 non-cooperative 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, non-cooperative 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 and a function that 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 postulate, not a limitation of the methodology. What is assumed to be exogenous in institutional analysis can be the object of institutional choice subsequently. Finally, the solution concept used is Nash equilibrium (possibly with a refinement) that 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.

For the study of institutions, non-cooperative game theory has two main advantages. 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 thus highlights *how* and *why* institutions matter.[15]

As an example consider the use of the Romer–Rosenthal agenda-setter model (Romer and Rosenthal, 1978, 1979) in legislative politics. 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 quite 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. The Romer–Rosenthal model's unique subgame perfect equilibrium outcome, however, predicts that the chosen policy will be biased away from

---

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

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. An institutionalist approach to the Romer–Rosenthal model, however, would not stop here but would compare different institutions in terms of their consequences. Baron (1996), for example, constructed a non-cooperative 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, non-cooperative game theory.

There is no need here to repeat the long and heated debate on whether models with or without such proposal power are the correct representation of, say, congressional decision-making.[16] After all, any model of proposal power needs to answer the question of why a chamber majority would ever grant such procedural prerogatives to some of its members. What is important in our context is the fact that this debate can be interpreted as a debate about the most appropriate game form to represent congressional decision-making. Different game forms are suggested, solved and then compared with empirical regularities. The fact that Nash equilibria exist in almost all applications has thus important consequences for the study of institutions, since it allows researchers to vary the institutional details by varying the game form and then compare the resulting equilibria.[17]

Game theory also allows an analysis of institutional choice, although explicit formal examples of equilibrium institutions are still rare. Calvert (1995) is among the most explicit adopters and 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 stating conditions under which deference to committees emerges endogenously. Thus, if one regards deference as an institution, it is an example of an equilibrium institution. Shepsle and

---

16. Summaries can be found in Krehbiel (1988) and (1991).

17. 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. True, the debate on proposal power has been mainly restricted to unidimensional 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.

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'. Diermeier and Myerson (1999) construct a bargaining model between chambers to explain cross-national differences in the internal organization of legislatures.

There are two properties of non-cooperative game theory that facilitate both institutionalist theories and theories of institutions. The first is the fact that equilibria exist under very general conditions. Second, games in extensive forms can be decomposed into subgames, where each subgame may correspond to a different institutional scenario. For example, in the Gilligan–Krehbiel model congressional decision-making under closed and open rules corresponds to two subgames under incomplete information with distinct informational and distributive properties. The choice of rules by the chamber can be modeled as a choice between two subgames determined by the floor median's expected equilibrium payoff. Thus, the institutional hierarchy discussed in steps A–E is formally captured as nested subgames of increasing complexity.

5. Revisiting Riker

Since its inception the New Institutionalism has been the subject of criticism, both within and outside of rational choice theory. As noted earlier, perhaps the most famous rational-choice-based critique is due to Riker (1980) who questioned 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 non-existence problems re-emerge. This problem became known as the *inheritability* problem. To quote Riker (1980: 445):

> In that sense rules or institutions are just more 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 also do institutions, except for short-run events.

From his 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. . . . (p. 443)

As we argued earlier, we agree with Riker that the inheritability problem presents a fatal blow to SIE-based institutionalism.[18] It does not follow, however, that there is a similar problem with the institutionalist research program as such. The problem with Riker's argument is that it tacitly identifies institutionalism with a specific methodology (SIE).[19] Riker and Shepsle both presuppose that the core is the adequate explanatory concept for a theory of politics. They disagree on whether SIE models suggest a solution to the non-existence problem.[20] But there is no need to rely on the core in institutionalist analyses. As discussed earlier, non-cooperative game theory does provide a competing explanatory concept that does not face similar existence problems. Non-cooperative game theory also allows us to construct theories of institutional choice without any apparent methodological problems.

Note that an institutional theory of politics abandons the very notion of an 'equilibrium of tastes', i.e. an 'institution-free' solution concept. This does not mean that the search for a general theory of politics must be abandoned as well or that political behavior is unpredictable. Rather, such a theory would have to be institutional at every level of analysis. A possible answer to Riker's view may thus state that the problem is not 'politics' but the insistence on an equilibrium concept that fails to generate any predictions in many political domains.

6. Conclusion

In this article we have developed an account of institutionalism as a methodology. The key idea of our interpretation is that there is no intrinsic difference between robust behavioral regularities and institutions, since the adoption and maintenance of institutions themselves are based on collective choice processes. This implies that institutions are best interpreted as theoretical constructs. Hence, which features of reality are to be treated as institutions amounts to a modeling decision that should be motivated by a theory's explanatory power. We argue that this approach allows researchers not only to compare political institutions but also to construct theories of institutions, i.e. 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.

---

18. See Diermeier (1997) for a more detailed discussion.
19. Of course, when Riker wrote his essay, SIE theory was the only game in town.
20. The notion 'equilibrium' in Riker's statement is ambiguous. Here it clearly refers to the core (equilibrium in the sense of General Equilibrium Theory).

We then argue that our interpretation 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 broad variety of institutions and institutional choice, strongly favors non-cooperative game theory over SIE theory. Third, Riker's famous inheritability argument does not present a problem for institutionalism but does undermine the use of SIE as the proper 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.

Baron, David (1996) 'A Dynamic Theory of Collective Goods Programs', *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,* pp. 216–28. New York: Cambridge University Press.

Cox, Gary W. (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. Stanford: 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* 89(5): 1182–96.

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.

Gilligan, Thomas W. and K. Krehbiel (1987) 'Collective Decision-Making and Standing Committees: An Informational Rationale for Restrictive Amendment Procedures', *Journal of Law, Economics, and Organization* 3: 287–335.

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

Groseclose, Timothy and James M. Snyder, Jr (1996) 'Buying Supermajorities', *American Political Science Review* 90(2): 303–15.

Hall, Peter A. and Rosemary C.R. Taylor (1996) 'Political Science and Three New Institutions', *Political Studies* XLIV: 936–57.

Kramer, Gerald H. (1972) 'Sophisticated Voting Over Multidimensional Choice Spaces', *Journal of Mathematical Sociology* 2: 165–80.

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: 873–90.

McKelvey, Richard D. (1976) 'Intransitivities in Multidimensional Voting Models and Some Implications for Agenda Control', *Journal of Economic Theory* 12: 472–82.

McKelvey, Richard D. (1979) 'General Conditions for Global Intransitivities in Formal Voting Models', *Econometrica* 47: 1086–112.

McKelvey, Richard D. and Raymond Riezman (1992) 'Seniority in Legislatures', *American Political Science Review* 86(4): 951–65.

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, CT: 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 Herbert Weisberg (ed.) *Political Science: The Science of Politics*. 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.

Shepsle, Kenneth W. and Barry Weingast (1994) 'Positive Theories of Congressional Institutions', *Legislative Studies Quarterly* 19(2): 149–80.

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.

DANIEL DIERMEIER is IBM Distinguished Professor of Regulation and Competitive Practice at the Kellogg School of Management at Northwestern University. His teaching and research has focused on government coalitions, legislative politics, formal theory, structural estimation, behavioral models and foundational issues in rational choice. ADDRESS: Department of Managerial Economics and Decision Sciences, Kellogg School of Mangement, Northwestern University, Evanston, IL 60208-2009, USA.
[email: d-diermeier@kellogg.northwestern.edu]

KEITH KREHBIEL is Edward B. Rust Professor of Political Science at the Graduate School of Business at Stanford University, where he has taught courses on foundations of political economy, legislative behavior, and business–government relations since 1986._ He specializes in American politics and has published two books and dozens of articles on U.S. politics and governmental processes. ADDRESS: Graduate School of Business, Stanford University, Stanford, CA 94305-5015, USA.
[email: krehbiel@stanford.edu]