Virtue and Self-Interest in the Design of Constitutional Institutions

*Lewis A. Kornhauser**

Constitutional political economy addresses four questions: (1) the causal question: What explains the constitutional institutions we observe? (2) the consequential question: What consequences do constitutional institutional have? (3) the ideal question: What constitutional institutions does justice require? and (4) the design question: What constitutional institutions are best for a polity given the constraints imposed by its current situation? Answers to the ideal and design questions require a theory of behavior that predicts how individuals will behave within constitutional institutions. Analysts usually assume that this theory of behavior corresponds to the explanatory theory developed to answer the second, consequential question. This essay argues that the assumption of rational self-interested behavior as the basis for a behavioral theory is not justified.

INTRODUCTION

Economic analysis of law recently extended its ambitions from the study and design of the institutions of private law to the study and design of the institutions of public law. In this project, which I shall call *constitutional political economy*,[1] one may identify four distinct questions. Two are explanatory, with the *causal* question addressing the problem of how we explain *which* constitutional institutions emerge in which countries;

---

* Alfred and Gail Engelberg Professor of Law, New York University. I thank Liam Murphy for comments on an earlier draft and the Filomen D'Agostino and Max E. Greenberg Research Fund of the NYU School of Law for financial support.

1 I thus transform this term from its current use as a label for a specific strain of economic analysis of political institutions — that associated with the work of James Buchanan — to a label for the general economic project. As I will note below, economic analysis of law encompasses at least two distinct traditions for the analysis of political institutions. One stems from the work of Buchanan and Tullock

and the *consequential* question addressing the problem of how we explain the effects of different constitutional institutions on the economic, political, and social development of a jurisdiction. Though controversy may rage concerning the details of these two inquiries, general agreement prevails at the level of methodology: conventional economic models of behavior should provide answers to these two questions.

The two remaining questions are not generally distinguished from one another, as analysis often conflates them into a single, normative question. Frequently, analysts pose the normative question in the form of: What constitutional structures ought we to have? This phrasing covers two distinct questions. The first question corresponds to the tradition in political philosophy that asks what constitution is best or what does justice require. I shall call this the *ideal* question. Two distinct approaches to the ideal question have emerged within constitutional political economy. I elaborate on the approaches of the welfarist and contractualist traditions below.

The second question provides a more pragmatic interpretation of the general question of what constitutional structures we ought to have. Rather than identify constitutional arrangements that are ideal under some (ill-) specified conditions, we might seek constitutional arrangements that are best for *us*, now, given current political conditions and the general predicament in which we find ourselves. With the fall of communism in Eastern Europe and the emergence of more democratic political institutions in parts of Asia, Africa, and Latin America, this *design* question has become particularly pressing and occupies increasing attention.

In this essay, I challenge the common understanding of the interrelations among these four questions. An outline of the common understanding thus sets the stage for my inquiry. The research program of constitutional political economy usually sees two important connections among these four elements of its research program. First, analysts adopt a single theory of how constitutional rules and institutions influence individual behavior. *Homo economicus*, the epitome of self-interested, rational action, serves not only as the route to explain the causes and consequences of constitutional institutions, but also to inform the ideal and design projects. Of course, the role of *homo economicus* in the ideal and design projects differs across the contractualist and welfarist strands of the normative enterprise; but in each, this role remains fundamental. Second, analysts have usually treated

---

and might be termed *contractualist*, while the other emerges from the literature on private law in economic analysis of law; it adopts a more *welfarist* approach in response to the normative question.

the design question as identical to, or derivative of, the ideal question. An answer to the ideal question largely resolves problems at the design stage.

This essay challenges both of these claims. I primarily contest an aspect of the first connection concerning *homo economicus*. At first glance, the theory of constitutional design appears to presuppose the explanatory theory. After all, a theory of design will itself have two components: an ideal criterion or objective at which the constitutional designer aims and a behavioral theory that projects how any set of constitutional institutions would function. The explanatory theory, it would seem, should serve as the behavioral component of the theory of constitutional design.[2] This natural elision of the explanatory theory and the behavioral component of the theory of design, however, has been largely resisted by leading constitutional political theorists.[3]

I evaluate and then largely reject the argument that the theory of constitutional design ought to adopt as a behavioral theory that of *homo economicus* even though *homo economicus* does not provide the best explanatory theory of constitutional behavior. The course of the argument leads me to a broader claim that severs the connection between the design and ideal questions. Specifically, I suggest that the pragmatics of design require a more eclectic and informal approach to the consequential question.

---

2 Some accounts of constitutional political economy would deny that the theory of design requires a behavioral component that must predict behavior of private individuals within specified institutions, because constitutional political economy does not select institutions on teleological grounds. It does not evaluate the institutions in terms of the actual consequences the institutions yield. *See, e.g.*, Geoffrey Brennan & James Buchanan, The Reason of Rules: Constitutional Political Economy 46-66 (1985).

3 *See, e.g.*, Geoffrey Brennan & James Buchanan, *The Normative Purpose of Economic 'Science' Rediscovery of an Eighteenth Century Method*, 1 Int'l Rev. L. & Econ. 155, 158 (1981) ("Simply put, our claim is that *homo economicus* rightly belongs in the analytical derivation of normative propositions about appropriate institutional design. In other words, the model of human behaviour that we might properly use in choosing among alternative institutions may be different from the model that would be more appropriate in making predictions about behaviour within existing institutional structures."); Dennis Mueller, Constitutional Democracy 50 (1996) ("As the underlying behavioral postulate for the purely positive analysis of all human behavior, the applicability of the rational self-interest assumption can be challenged. But as the foundation of a *normative* analysis of political institutions, the assumption seems unassailable. Even if the individuals do not always act in perfectly rational ways, they would presumably wish to be governed by those institutions that they would rationally choose."). As I will mention further in the notes below, Brennan's view has evolved to one that accords more with the view I offer in this essay. *See* Geoffrey Brennan & Alan Hamlin, Democratic Devices and Desires (2000).

I. REFINING THE NORMATIVE QUESTION

The normative question asks: What constitutional institutions ought we to have? Obviously, the answer to this question will depend on one's conception of justice and on one's understanding of the nature and function of government. Indeed, debates over the normative question often concentrate on this disagreement; as I will develop later, the dispute between the welfarist and contractualist schools of welfare economics is formulated around this disagreement.

The controversy over the appropriate conception of justice, however, has masked an important equivocation in the question: What institutions we ought to have depends critically on the formulation of the circumstances in which the institutions must operate and the circumstances under which the agents must choose the institutions. The circumstances of operation and choice may each be further divided into what I shall call *environmental* circumstances and *motivational* circumstances. Environmental circumstances specify the environment in which agents must act; that is, they specify the environment in which agents act *within* on-going constitutional institutions and the environment in which agents *choose* the constitutional institutions. Similarly, the motivational circumstances refer to the motivations of the agents in each context. Notice that I have identified two sets of environmental and motivational circumstances: the choice set, which applies at the time constitutional institutions are chosen, and the operational set, which applies within on-going constitutional institutions. Of course, these two contexts are not always neatly differentiated; at least occasionally — but, arguably, all the time — agents may alter the constitutional institutions within which they are operating.

Though the literature often fails to distinguish among them, the normative question has many variants. I distinguish two important variants, the ideal question and the design question. These two questions adopt polar specifications of the circumstances under which agents must choose institutions (the *choice circumstances*). Not only do the specifications of the choice environments differ, but the specifications of the motivations of the agents differ dramatically as well.

The tendency to conflate a number of distinct questions within the normative question is not restricted to economic analyses of constitutional arrangements. It will facilitate discussion, however, to investigate this conflation in the context of the two distinct traditions within welfare economics that address the normative question. The first subsection below thus briefly outlines these two traditions. Subsequent subsections specify in

more detail the two variants of the normative question under consideration: the ideal and design questions.

A. Two Traditions of Economic Evaluation

Economic analysis of law draws on two distinct traditions to answer the abstract, normative question: What institutions ought we to have? I call these the *welfarist* and the *contractualist* traditions.

The welfarist tradition, represented by such authors as Arrow and Samuelson,[4] assesses constitutional institutions solely in terms of the well-being of individuals. The focus on well-being leaves much open to controversy and elaboration. Most obviously, one must specify a conception of well-being that is both ethically attractive and operational (or measurable). Many further questions arise as well. One must, for example, determine *whose* well-being is relevant.[5] Moreover, one must determine precisely how well-being is to be aggregated across individuals. A welfarist need not ignore distributional concerns.[6] For purposes of defining the contrast with the contractualist tradition, however, the welfarist commitment to evaluation of institutions in terms of the results they induce is central.

The contractualist tradition assesses constitutional institutions in terms of the consent of the individuals governed by those institutions. This tradition denies that one can evaluate constitutional regimes in terms of the end-states to which they lead.[7] Rather, the institutions must be evaluated directly. Direct evaluation generally is formulated in terms of *consent*. Consent focuses on each individual's assessment of the institutions. Moreover, this evaluation is generally assumed not to rely solely on the consequences of the institutions

---

4 *See* Kenneth J. Arrow, Social Choice and Individual Value (2d ed. 1963); Paul A. Samuelson, Foundations of Economic Analysis 203-53 (enlarged ed. 1983). Welfarism is the dominant tradition in welfare economics, and it is manifest in the attention given to cost-benefit analysis.

5 Generally, the assessment considers only the well-being of individuals *within* the jurisdiction governed by the constitutional institutions. The restriction to this set of individuals requires some defense if one seeks to ground the welfarist tradition in some theory of justice. Welfarist assumptions seem to dictate attention to the well-being of everyone, regardless of whether they are subject to the jurisdiction.

6 Louis Kaplow & Steven Shavell, *Fairness vs. Welfare*, 114 Harv. L. Rev. 961 (2001), for example, emphasize this in their general defense of welfarism.

7 Brennan & Buchanan, *supra* note 2, argue against end-state evaluation on somewhat different grounds than those offered in the text. They suggest that the subjective nature of individual well-being prevents end-state evaluation. *Id.* at 45.

for the individual. Complications of course arise in elaborating an appropriate conception of consent that does not look to end-states.[8]

B. The Motivational Circumstances

Under some theories of justice, the analyst ascribes motivations to actors at the time constitutional institutions are adopted that are radically different from the motivations ascribed to actors inhabiting those institutions. The ideal and design questions differ radically in the motivations that they ascribe at the time of adoption.

So, for example, in Rawls' theory of justice, agents, at the constitutional moment — behind the veil of ignorance — are motivated by justice. Once the agents assume positions within the on-going, chosen institutions, however, their motivations are more complex and often tainted. Self-interest rather than justice may dictate in part their choices. Notice that the ideally motivated agents who *adopt* constitutional institutions may know, as in Rawls, the non-ideal motivations of the agents who will *inhabit* the institutions. The ideally motivated agents choose the constitutional institutions in light of their knowledge of the non-ideal actors.

This dual motivational structure has been criticized as incoherent.[9] It does reflect, however, two distinct concerns that lie at the heart of the normative question. One concern arises out of the need for institutions to function in the real world. One cannot presume that individuals will conform their behavior to institutional requirements. Rather, institutions must be, as economists say, *incentive compatible*; actual behavior depends on institutional rules and agent motivations. Incentive compatibility thus constitutes a constraint on

---

8 Thomas M. Scanlon, What We Owe to Each Other (1998), provides an analysis of reasonableness within the contractualist tradition.

As will be developed later, the contractualist tradition in welfare economics links consent and preference. Individual preferences, however, are generally over end-states or consequences. To equate consent and preference by saying that X consents to institutions A over set B if and only if X prefers the outcomes of A to the outcomes of B does not avoid the evaluation of end-states. Phrased differently, the contractualist tradition apparently asserts contradictory claims: on the one hand, that consent to constitutional institutions does not involve evaluation of end-states; and on the other hand, that preferences over end-states provide the criterion for consent.

9 *See* Gerald A. Cohen, *Where the Action Is: On the Site of Distributive Justice*, 26 Phil. & Pub. Aff. 3 (1997); Gerald A. Cohen, *The Pareto Argument for Inequality*, 12 Soc. Phil. & Pol'y 160 (1995); Liam Murphy, *Institutions and the Demands of Justice*, 27 Phil. & Pub. Aff. 251 (1999).

the structure of constitutional institutions. The second concern has a more normative cast. Analysis must specify the grounds on which each agent accepts or endorses constitutional institutions.

Generally, to answer the design question, the analyst assumes that the parties who must adopt the constitutional institutions have the actual motivations they do have at the time they make the design decision. As actual design occurs while agents inhabit some set of institutions, analysis of the design question naturally assumes a motivational symmetry between the agents who design the institutions and the agents who will inhabit them. This assumption contrasts with that made when investigating the ideal question; there the analyst generally assumes that the agent's motivations have been purified or abstracted in some way at the time the constitutional institutions are chosen.[10]

The concern with actual motivations in analysis of the design question arises because any proposed constitutional institutions must be acceptable to the parties drafting the constitution. Normative conceptions of consent are irrelevant in a context in which the institutions must function for and with the agents who adopt them. The design of the U.S. Constitution, for example, reflects this concern at several points. Both small states and slave states had an interest in the existence of a bicameral legislature in which representation in one house did not depend on population. Ideal constitutional institutions might well not be bicameral. Similarly, the provisions expressly dealing with slavery respond directly to the concerns of slave states. Many constitutions in the recent round of constitution-drafting, spurred by the transition to democracy in various countries, exhibit similar qualities.

In the analysis of the ideal question, by contrast, motivations are assumed to be ideal in some sense. The welfarist assumes ideally "altruistic" motivations; or, phrased differently, she assumes an impartial or impersonal acceptance criterion. Each individual seeks to maximize the social welfare function, so each pursues the set of institutions that will accomplish the shared end. The contractualist assumes a diluted form of self-interested rationality. The dilution occurs through the specification of the environment in which each agent is assumed to choose.

---

10 The contractualist tradition has criticized the welfarist tradition for its asymmetric assumptions about public and private actors. In the context here, the criticism is directed at the assumption in the ideal stage that motivations at the time of constitutional choice differ from the motivations the agents will have within the institutions.

C. The Environmental Circumstances

The circumstances in which individuals make choices also point to different questions within the rubric "normative." Again, it may be useful to distinguish two types of environmental circumstances. The first concerns the uncertainty that agents at the acceptance stage face concerning their situation under the constitutional institutions. The second type concerns the uncertainty that the agents face concerning the range of situations under which the constitutional institutions will operate.

Consider first uncertainty concerning the agent's own situation. Analyses of the design question generally assume that the agent faces significantly less uncertainty than she does under analyses of the ideal question. One might say that from the design perspective, one assesses constitutional institutions in terms of their performance or other characteristics relative to the *actual* population on the assumption that each agent knows (at least her current) situation within the social structure. From the ideal perspective, by contrast, analysts assess constitutional institutions in terms of their performance relative to a broader set of population profiles in which each agent is uncertain of her situation in the social structure.

This characterization points to two differences in the analytic stance that correspond to important aspects of Rawls' analysis in *A Theory of Justice*[11] and Kenneth Arrow's analysis in *Social Choice and Individual Values*.[12] The social choice tradition seeks social decision rules that satisfy specified criteria, generally including one of universal domain. This condition requires the social decision rule to apply to *any* population, regardless of the distribution and content of preferences or values of the population. In the more concrete context of the design of constitutional institutions, universal domain seeks constitutional institutions that would be best or effective in any polity; constitutional institutions must work in Sweden, the United States, and India among other countries. This condition may be implicit in analyses of the ideal question, but it is alien to the design perspective seeking constitutional institutions that need succeed only in the polity that accepts and adopts them.

It is worth noting that in the contractualist tradition, the agent's uncertainties about her position in the social structure and the polity to which the constitutional institutions will apply serve to transform the self-interested agents who will inhabit the constitutional institutions into the more impartial, perhaps altruistic individuals who must accept and adopt

---

11 John Rawls, A Theory of Justice (2d ed. 1999).

12 Arrow, *supra* note 4.

them. Moreover, the normative force of the consent arguments underlying the contractualist position is unclear when the argument is transferred from the ideal to the design context.

A similar contrast in the relative generality of the two perspectives persists when one turns to consideration of the set of economic, political, and social circumstances under which the constitutional institutions must operate. As in the earlier case, the design perspective adopts a much narrower conception of the set of operational circumstances: the constitutional institutions must work, primarily, in the specific polity in its particular circumstances. Consider, for example, polities such as South Africa or Indonesia, which include significant ethnic divisions with respect to income, wealth, educational background, and social position. The design of a constitutional framework must promise success primarily within the given ethnic divisions, even if the polity aspires to a less-divided social life and organization.[13]

Investigations of the ideal perspective, by contrast, often leave the environmental circumstances unspecified or specified only in a general sense. A distinction, for example, is often made between "lifeboat circumstances" and other, more favorable ones. Rawls' theory of justice, for example, is meant to apply only to polities that are above subsistence.

II. The Relation between Ideal, Design, and Behavioral Theories

Before investigating the claim that one ought to use *homo economicus* as a behavioral theory to inform one's answers to the ideal and design questions, I must first confirm that answers to these questions require use of *some* behavioral theory. As behavioral theories have a different status in the welfarist and contractualist traditions, the argument must here bifurcate. I begin with a discussion of the welfarist tradition, where the importance of a behavioral theory is most obvious.

A. The Welfarist Tradition

The importance of a behavioral theory to a resolution of the ideal and design questions within the welfarist tradition derives immediately from the

13 In some instances, a "transitional" constitution is drafted, but it is noteworthy that such transitional documents often persist.

instrumental role of constitutional institutions within the welfarist tradition. The welfarist tradition answers the ideal and design questions in a similar way: choose those constitutional institutions that maximize the social welfare function under the relevant conditions.[14] Determining which constitutional institutions optimize the social welfare function requires one to predict social welfare under each set of institutions. Prediction of social welfare in turn entails the prediction of how individuals will behave under the institutions.

However, at least in the context of the ideal question, this predictive task seems incredibly complex for several, related reasons. First, constitutional effects on social welfare are indirect. Constitutions generally determine the structures in which public officials will act; the actions of public officials then determine the structures in which private citizens will act. The complex combination of choices of public officials and private individuals then determines the well-being of each individual.

Second, and related to the first, determination of the ideal constitution may occur in a context in which the preferences of the citizenry to be served may be uncertain. In this case, the constitutional analyst seeks the regime that will maximize expected social welfare. This determination, however, is also complex. After all, the set of constitutional institutions that maximizes expected social welfare need not be a set of institutions that maximizes social welfare for one of the possible preference profiles of a population.[15]

Third, a polity may confront a diverse set of economic, social, and political conditions. A different constitutional regime might best promote social welfare under each of the possible set of conditions. So, for example, institutions that serve a polity well in times of abundance and peace may ill-serve the polity during times of war and famine. However, if the designer seeks institutions that maximize expected social welfare, she cannot restrict her attention to institutions that are optimal only under one set of conditions or another. A set of institutions that is both second-best during times of peace and second-best during times of war may perform better, on average, in the uncertain world. Moreover, the analytic problem is further complicated by the fact that constitutional regimes may affect the likelihood that conditions of war or peace, famine or abundance may arise. Optimal choice requires attention to these feedback effects.

---

14 Thus, the two questions differ only in the set of constraints against which maximization of the social welfare function occurs. In the design question, the constraint set is larger.

15 I have assumed that a welfarist designer seeks to maximize expected social welfare, but as I suggest later, she might have a different goal that still retains a welfarist flavor.

The predictive task is somewhat easier in the analysis of the design question because the designer faces fewer uncertainties than the philosopher. Nevertheless, because constitutions induce behavior only indirectly, the predictive task remains formidable, perhaps sufficiently formidable to call the enterprise into question. Indeed, the contractualist tradition in welfare economics in part arises as a response to the difficulty of the predictive task.

B. The Contractualist Tradition

Consent rather than well-being serves as the key evaluative concept in the contractualist tradition. This tradition thus appears to eschew an instrumental account of the evaluation of constitutional institutions. Moreover, a non-instrumental account of evaluation apparently does not require a theory of behavior.

In the contractualist tradition, behavior reenters the evaluative project because the tradition elaborates its concept of consent in terms of preference. More specifically, the contractualists apparently argue that an agent will consent to X over Y if and only if she prefers X to Y. This conception of consent presents several difficulties that I largely ignore here in favor of pursuing the question of how this strategy requires the contractualists to employ a theory of how individuals behave.[16]

To apply this conception of consent to the problem of the choice of constitutional institutions, one must identify the domain over which individuals have fundamental (or primitive) preferences. Unfortunately, it seems unlikely that individuals have preferences directly (or basically) over constitutional regimes rather than preferences over regimes that derive from more fundamental preferences over outcomes. If, however, agents' preferences over constitutional institutions derive from preferences over outcomes, one must predict what outcomes each set of constitutional

16 Nonetheless, I note in passing a few difficulties with a conception of consent that relies on preference. First, consent is a normative notion. The contractualist must thus defend why consent should be equated with (rational) preference. Such a defense is difficult because: (a) not all accounts of rationality rest on preference; and (b) in many instances, we think that an agent has consented even when she has not acted in a (fully) rational way. Second, the equation of consent with preference may yield only an incomplete criterion in the context of choice over constitutional institutions, because the feasible set of institutions contains more than two options and agents may differ in their preferences over these alternatives. So, though both R and S may prefer X and Y to Z, R may prefer X to Y while S prefers Y to X. Consequently, R and S cannot agree on a set of constitutional institutions.

institutions will induce. One thus confronts the same difficulty faced in the welfarist tradition.

Brennan and Buchanan suggest that this problem is overcome when self-interested individuals choose in the context of uncertainty over which positions they will occupy in society. Under these choice conditions, they argue, self-interest dictates the choice of institutions "that eliminate or minimize prospects for potentially disastrous results."[17] The argument for this conclusion relies on the context of uncertainty, which implies that each agent adopts a more impartial attitude in her evaluation of institutions because she does not know which position in society she will occupy.

This argument for the outcome of choice under uncertainty has several gaps. First, Brennan and Buchanan do not show that in fact, optimal choice in these conditions is given by a maximin criterion (as their formulation suggests). Second, it is not clear why choice by *self-interested* individuals should provide the normative baseline for the evaluation of constitutional institutions.

Moreover, the Brennan and Buchanan argument does not apply to the design question, because constitutional design does not occur in a context in which individuals do not know which positions they will occupy in society. Individuals will rely on actual preferences in deciding whether to adopt a set of constitutional institutions. Since in any real context of constitutional choice, the agents who design and choose a set of constitutional institutions will also be the agents who inhabit those institutions, real problems of constitutional choice are questions of design theory, not ideal theory.

III. *Homo Economicus* in an Ideal Theory

The argument of the previous section established that to address normative questions, both the welfarist and contractualist traditions require a theory of how individuals will behave within the constitutional institutions that they select. For the welfarist tradition, this requirement emerges directly from its instrumental approach to the analysis of institutions. For the contractualist tradition, the theory of behavior arises indirectly from a theory of consent that ties consent directly to preference.

In the next two sections, I discuss the choice of a behavioral theory to address the ideal and the design questions. As the literature does not generally

---

17 Brennan & Buchanan, *supra* note 2, at 29.

distinguish between these two questions, the division of the discussion is somewhat arbitrary.

Dennis Mueller has argued that the rationality assumptions embodied in *homo economicus* provide the appropriate basis for the evaluation of constitutional institutions:

> As the underlying behavioral postulate for the purely positive analysis of all human behavior, the applicability of the rational self-interest assumption can be challenged. But as the foundation of a *normative* analysis of political institutions, the assumption seems unassailable. Even if the individuals do not always act in perfectly rational ways, they would presumably wish to be governed by those institutions that they would rationally choose.[18]

One might interpret this argument within either the welfarist or contractualist traditions of welfare economics. Unfortunately, it does not succeed in either tradition.

Consider first the welfarist tradition. In this tradition, the behavioral theory directly predicts how individuals will behave within the constitutional institutions. Recall that in the tradition, the polity is agreed that its constitutional institutions should maximize social welfare and that they share a conception of social welfare. Why would they estimate the consequences for social welfare of each set of constitutional institutions by using a theory of behavior — *homo economicus* or any other — that they know to be false? In the next section, I shall consider, and reject, several possible answers to this question in the design context. Here it suffices to note that Mueller's claim is certainly assailable. Why should institutions that would be best if everyone were rational recommend themselves to individuals who know that they themselves — the future inhabitants of those institutions — are less than fully rational? Under the welfarist tradition, the individuals choosing constitutional institutions are seeking to maximize social welfare for the individuals who will inhabit those institutions. To ignore systematic irrationality on the part of the inhabitants of the institutions will lead the designer to miscalculate the consequences for social welfare of each set of institutions.

Turn now to the contractualist account, which probably accords better with Mueller's own position. In this tradition, the rationality assumption might apply to either the agents at the time of choice or to the agents who will

---

18 Mueller, *supra* note 3, at 50.

inhabit the institutions. Presumably, Mueller intends the choosing agents to be fully rational, but the inhabiting agents to retain their imperfectly rational motivations.[19] The choosing agents would then rationally select the set of institutions that would be best for the actual agents.

Let C* be the set of constitutional institutions to which fully rational individuals would consent (or, equivalently on the contractualist account, would prefer). Suppose individuals are not fully rational. What recommends the constitutional institutions to which fully rational agents would consent to these imperfect agents? Presumably, they would not choose them. Nor would they obviously consent to them.[20] Moreover, even a fully rational individual might not prefer the set of institutions C* to some other set of constitutional institutions that are adapted to the cognitive processes of the polity. Institutions adapted to the cognitive errors of the polity are apt to function better and, hence, better promote the well-being of all individuals governed by the institutions.

IV. *HOMO ECONOMICUS* AS THE BEHAVIORAL COMPONENT OF A THEORY OF CONSTITUTIONAL DESIGN

As noted earlier, various theorists have denied the suitability of the explanatory theory as the behavioral component of a theory of constitutional design. Geoffrey Brennan and James Buchanan have been the leading advocates of this disjunction between explanatory and normative theories.[21] To begin, then, I set out their argument in some detail in Part A. In subsequent parts, I argue that two of the three steps in their argument are open to question.

---

19 If the inhabiting agents are also assumed to be fully rational, then the argument used in the welfarist tradition applies.

20 Consent is a moral notion, not an explanatory one. Terms of consent identify the circumstances under which choice has moral significance. Of course, one might argue that choice only has moral significance when the agent acts in a fully rational fashion. This argument, however, would exclude most market transactions from the realm of moral acceptability.

21 *See* Brennan & Buchanan, *supra* note 2; Brennan & Buchanan, *supra* note 3. Geoffrey Brennan has since rejected the position for which he and Buchanan have argued. *See* Brennan & Hamlin, *supra* note 3.

A. The Argument of Brennan and Buchanan[22]

Brennan and Buchanan mark the distinction made earlier between an explanatory theory and the behavioral component of a theory of design as a distinction between an explanation of behavior of agents within given institutions and a comparison of alternative institutions. They seek to defend the use of *homo economicus* in the comparison of alternative institutions; that is, they seek to defend *homo economicus* as the behavioral component of a theory of design. By *homo economicus*, they mean an economic theory of behavior narrowly understood. Thus, each agent seeks to maximize wealth.

Their argument for *homo economicus* has three steps.

First, they argue that the primary theoretical need in the comparison of alternative institutions is the use of a uniform theory of behavior. That is, one must use the same behavioral theory to evaluate each institution under consideration. Brennan and Buchanan offer a methodological justification for this requirement. They claim that unless one adopts a uniform theory of behavior for comparison of institutions, one will be unable to attribute differences in behavior to differences in institutions.[23]

Second, they argue that in the comparison of institutions, one need not use the theory that best predicts behavior. Indeed, their argument suggests that one should *not* use the best predictive theory of behavior in design, because the constitutional designer should be risk-averse relative to extremely adverse outcomes.[24] In what follows, I do not challenge this claim, though I do note that it requires defense.

Third, they argue that *homo economicus* is the appropriate uniform theory to use in the comparison of institutions. They offer four reasons for this conclusion. (1) They suggest, following Hume and Kant,[25] that political institutions should be designed on the assumption that they will be run by knaves; that is, one should design institutions on the assumption of

---

22 I rely primarily on Brennan & Buchanan, *supra* note 3. Brennan, in collaboration with Alan Hamlin, subsequently modified his position, Brennan & Hamlin, *supra* note 3. I shall note some of these changes where appropriate.

23 Brennan & Buchanan, *supra* note 3, at 159.

24 Brennan & Buchanan, *supra* note 2, at 55.

25 I David Hume, *Of the Independence of Parliament, in* Essays, Moral and Political (1742), *reprinted in* I David Hume, Essays: Moral, Political and Literary 42-46 (E.F. Miller ed., 1987) ("Political writers have established it as a maxim, that, in contriving any system of government, and fixing the several checks and controls of the constitution, every man ought to be supposed a *knave*, and to have no other end, in all his actions, than private interest."); Immanuel Kant, Eternal Peace (1795).

worst behavior. (2) Institutions should be designed to transform self-interested behavior into publicly desirable behavior, because we should expect some self-interested officials within government. (3) In support of argument (2), Brennan and Buchanan argue that a Gresham's law of motivation applies: self-interest motivations drive out other motivations, as non-self-interested individuals need to defend themselves against the self-interested. (4) In a less extreme argument than (1), Brennan and Buchanan contend that social harm is a convex function of the deviation of behavior from optimal behavior.

In the following Parts, I address the uniformity claim and each of the four arguments Brennan and Buchanan offer in support of the claim that *homo economicus* is the appropriate uniform theory of behavior.

B. The Need for a Uniform Theory

The use of a uniform theory of behavior to compare the functioning of alternative institutions will be appropriate only if the motivations of actors within institutions are independent of the institutions. The assumption that motivations are exogenous, however, seems implausible on at least two counts.

First, different institutions might attract different motivational types. Different monastic orders, for example, presumably attract adherents with different preferences and motivations. Franciscans who accept vows of poverty presumably attract acolytes with different tastes than Jesuits. One might argue, of course, that the selection mechanism that explains the attraction of different types must rely on some underlying, uniform theory of behavior that explains why individuals are attracted to institutions of particular types.

Second, different institutions might shape individual preferences and motivations differently. The shaping of preferences, of course, refers to a phenomenon distinct from the inducement of different behaviors, because of differences in incentive structures. An individual with fixed preferences may behave differently under different incentives. Some institutions, however, may alter the preferences that an individual actually has. As we have little understanding of the process of preference formation, it is often easier to ignore the problem than to analyze it.

The claim that comparison of institutions requires the use of a single, uniform theory faces another difficulty as well. Suppose one compares institution A to institution B under the assumption of narrowly self-interested behavior and discovers that A is preferable to B. Now, suppose that one performs the same comparison on the assumption of altruistic behavior (or of less narrowly self-interested behavior) and discovers that institution B

is preferable to institution A. The choice between A and B reduces to a judgment concerning the applicability of the behavioral theories. Thus, the choice of *which* behavioral theory to use in the comparison is crucial. I now turn to arguments concerning that choice.

C. *Homo Economicus* as the Theory of Worst Behavior[26]

A designer might adopt one of at least four different approaches to the determination of the behavioral consequences of a given legal rule or institution. First, she might use the *best predictive theory of behavior* to predict the consequences of the rule or institution. That is, she identifies the theory that best predicts how private individuals and, if necessary, public officials will behave; she then uses that theory to predict how the rule or institution will function. Second, she might adopt a more optimistic approach in her assessment and use the *theory of best behavior*. The designer thus imputes to the relevant agents the motivations that insure that the rule or institution being designed functions best from the designer's point of view. Third, the designer might adopt the *theory of worst behavior*; i.e., the designer imputes to the agents the motivations under which the rule or institution functions *least* well. Fourth, the designer might seek the *best prediction of the behavioral consequences* of the rule or institution. If the designer were Bayesian, on this account, she would assess the rule or institution in accordance with each theory of behavior and then weight the assessments by her beliefs about the likelihood that the given theory of behavior will accurately describe the consequences of the rule or institution. She need not, however, be a Bayesian and need not base her predictions on any *theory* at all. So, for example, in the formulation of tax policy, the analyst might predict the consequences of a new tax on the basis of past behavior, but without any theory that predicts individual behavior.[27]

These theories may not always be distinct. One might believe, for instance, that in some circumstances, the theory of worst behavior is also the best predictive theory of behavior. One might believe this, for example, about the economic theory of behavior in general or in a particular context.[28]

The last comment suggests that each of these theories might be specific

---

26 This section draws on Section 4 of Lewis A. Kornhauser, *Three Roles for a Theory of Behavior in a Theory of Law*, 31 Rechtstheorie 197 (2000).

27 I thank Ed McCaffrey for suggesting the example of prediction of the effects of tax laws.

28 In fact, the economic theory of behavior does not coincide with the theory of worst behavior. Malevolent actors who seek to maximize harm to others would surely lead

to the legal rule or institution under examination. One might believe, for instance, that to evaluate legal rules that govern transactions between merchants, the economic theory of behavior is the best predictive theory. In contrast, one might believe that to evaluate consumer behavior, the best predictive theory incorporates various aspects of psychological theories that account for systematic errors in rationality made by consumers. Finally, one might believe that an obligation theory is the best predictive theory for the conduct of public officials when considering a rule that modifies an ongoing public institution. Indeed, in the evaluation of the rule's effects on official conduct, the designer might believe that the economic theory of behavior is the theory of *worst* behavior.

One might argue similarly that the theory of best behavior, as well as the theory of worst behavior, will be rule- and institution-specific. *A fortiori*, then, the beliefs of a Bayesian designer regarding the likelihood that any particular theory predicts accurately will depend on the rule or institution under assessment.

The dependence of the best theory of behavior on the rule or institution under consideration arises because the motivations of the individuals governed by the rule or the public officials who staff the institution are in part endogenous. This endogeneity might arise in two ways. The legal rule might act directly on the motivational preferences of the private individuals or public officials so that they come to have different motivations than those they would have had in the absence of the rule or institution. Alternatively, the legal rule might act indirectly through some selection mechanism. Those private individuals or public officials with a particular type of motivation might be differentially attracted to work in an institution with a specific design or to engage in transactions governed in a particular way.

The endogeneity of motivations is clearest in the context of the motivations of public officials. The structure of the institutions may determine in part who becomes a public official. The nature and amount of remuneration, the appointment (or election) process, and the allocation of power across posts and institutions will influence who will choose to contest and win elections and who will serve in various bureaucratic posts. A patronage system, for example, will yield a very different public administration than a civil service system will yield. Put differently, a motivational assumption may be self-fulfilling. Such a result would be particularly disheartening if one were to adopt a worst-case scenario as one's predictive theory.

---

to worse results. Even arbitrary or unpredictable actors would likely lead to worse results than self-interested ones.

One might argue that the possibility of selection simply moves the question of theory adoption back one step from the choice of a theory that explains behavior within an institution to the choice of a theory that explains which individuals will select into an institution. This argument clearly holds for a Bayesian, who would determine, under each theory, the distribution of motivations that public officials would have and then evaluate the institution on the basis of these beliefs about the likelihoods of differently motivated compositions of the institutional staff. For the other theories, however, the problem of selection does not simply push the question one step back, because in general, no selection mechanism will be perfect. Consequently, an institution will always include officials with a variety of motivations.

D. Self-Interested Behavior Drives out Other Behaviors

Brennan and Buchanan, in what might be understood as a corollary to the argument that it is the worst behavior, also defend the use of *homo economicus* as the behavioral component of the theory of design on the grounds that self-interested behavior "drives out" other types of behavior.[29] This argument has several ambiguities as well as several gaps.

To begin, consider the ambiguities. First, the range of behaviors driven out by self-interest requires specification. Are only benevolent or altruistic preferences driven out? Or does self-interest undermine malevolent preferences as well?

Second, the argument does not specify whether non-self-interested behavior is driven out of society altogether or only from the public sector.

Third, the mechanism by which self-interested behavior supercedes other behavior is not identified. Three distinct processes might be at work, singly or in combination. In one process, behaviors might be filtered, screened, or sorted so that self-interested behaviors appear in one social forum but not in others. According to a second process, social processes might select for self-interested behavior so that non-self-interested behavior becomes extinct. Some economists argue, for example, that markets select for profit-maximizing producers because producers that do not maximize profit would be driven into bankruptcy.[30] Under a third process, non-self-interested individuals might be converted to self-interest.

Consider the screening process argument first. Brennan and Buchanan

---

29 Brennan & Buchanan, *supra* note 2.

30 Prajit K. Dutta & Roy Radner, *Profit Maximization and the Market Selection Hypothesis*, 66 Rev. Econ. Stud. 769 (1999), argue that this selection claim is false.

suggest that self-interested individuals will have an advantage in politics; consequently, the political process will screen for self-interested individuals, so that political positions will be filled disproportionately (or completely) by self-interested agents.

Consider now selection mechanisms. Notice that whether social mechanisms select for self-interested behaviors will depend on the details of the social mechanism. In simple models, some cooperative (or partially cooperative) behaviors will survive in an environment of self-interested actors or, in some circumstances, drive the self-interested behavior to extinction.[31] Indeed, even contexts in which selection for self-interested behavior seems intuitive may not, in fact, select for such behaviors. Consider profit-maximizing behaviors in markets. In some models, profit-maximizing firms are certain to fail in finite time, while some non-profit-maximizing strategies can earn positive expected profits and have a positive probability of surviving forever.[32]

Finally, though non-self-interested individuals might be converted to self-interested ones, the process of conversion requires specification in order for the argument to be compelling. Intuition does not suggest that political institutions are the ones likely to convert agents within them to self-interested behavior.

In sum, we have little reason to believe that self-interested behavior will drive out other behaviors, even in market institutions. It seems unlikely, then, that self-interest will drive out other motivations in political institutions.

E. Economizing on Virtue (as a Refinement of the Study of Markets)

Constitutional political economy heavily emphasizes the study of institutions that transform the self-interested behavior of individuals into outcomes that serve the interests of all (or at least of others). The market, of course, is the paradigmatic institution to exhibit this property. I shall call institutions that successfully convert self-interested behavior into socially optimal behavior *incentive-based institutions.*

The investigation of how institutions that transform self-interested behavior into outcomes that serve the interest of all is an important and difficult task. It is not clear, however, why this concern should dominate the choice of behavioral theory to be used in answering normative questions.

---

31 *See, e.g.*, Robert Axelrod & William D. Hamilton, *The Evolution of Cooperation*, 211 Sci. 379 (1981).

32 Dutta & Radner, *supra* note 30.

Discussions suggest two related reasons for the emphasis on incentive-based institutions. First, one might argue that even if not all agents act self-interestedly all of the time, self-interested behaviors are more reliable; consequently, one should "economize on virtue" and design institutions that function well when agents act self-interestedly. Second, and related, the argument implicitly assumes that incentive-based institutions will also work well when agents are virtuous. If this assumption is false, the argument for the use of incentive-based institutions fails. If incentive-based institutions perform badly when inhabited by virtuous agents, a designer who is uncertain about the motivations of the agents who will inhabit the institutions may wish to select institutions that are more robust to various agent motivations. Alternatively, the designer might shift attention to the design of mechanisms that more successfully select for virtuous agents.

In this Part, I argue the fallaciousness of the claim that the use of incentive-based institutions is innocuous because if an institution performs well when populated by self-interested individuals, it will perform well when populated by individuals with other motivations, particularly more salutary ones.[33]

Consider directly the implicit claim that an institution that performs well when agents have self-interested motivations will perform well when agents have other motivations. Moreover, consider Adam Smith's classic discussion that it is "not from the benevolence of the butcher that meat gets on the table." Market institutions that are populated with altruistic agents may not work as well as they would if all agents were self-interested. If the butcher gives away meat or underprices it, then the market prices will no longer provide accurate signals of the opportunity costs of resources and the resulting equilibrium may no longer be Pareto-optimal. Phrased more abstractly, altruistic motivations violate the assumption on convexity of preferences that underlies the theorems guaranteeing a competitive equilibrium and that every competitive equilibrium is Pareto-optimal.[34]

A number of other examples show that an institution may perform better if agents do not have purely self-interested motivations. To understand the argument requires that one distinguish between the game form and the game itself. The game form specifies the players, their moves (or strategy sets), and the outcomes, while the game itself specifies the players, their

---

33 *See, e.g.*, Brennan & Hamlin, *supra* note 3, at 299 ("All that is required for such results is the assumption that agents are concerned about their own interests *among other things*.").

34 For a statement of these theorems, see Gerard Debreu, A Theory of Value (1959).

moves, and the payoffs. That is, the game form is transformed into the game when each player's preferences over outcomes are specified. One may thus identify an institution with a game form and the game with the incentive structure faced by specific individuals placed in institutional roles.

Both the distinction between game form and game and the role of motivations in determining outcomes may be illustrated by an analysis of the classic prisoner's dilemma. To begin, we must specify the game form underlying the standard normal presentation of the prisoner's dilemma. Table 1 presents the institution of plea-bargaining. The plea-bargaining *institution* is defined by the matrix of sentences incurred by the defendants as a function of their decisions to remain silent (cooperate) or to confess (defect).

Table 1
The Institutional Structure of the Prisoner's Dilemma: Sentences as a Function of Action

ROW/COLUMN	Cooperate Remain Silent	Defect Confess
Cooperate/Remain Silent	6 months, 6 months	5 years, probation
Defect/Confess	probation, 5 years	3 years, 3 years

Table 1 presents a game form or institutional structure that underlies the much-analyzed game of the prisoner's dilemma. The matrix identifies what "physical" outcome results when each player adopts a strategy. The first number in each cell represents the sentence that Row will receive if each player undertakes the corresponding strategies. The second number in each cell represents the sentence that Column will receive. Without specifying the preferences of the players, one cannot determine how the institution will perform. Phrased differently, a game is defined by the game form or institutional structure and the preferences of the agents who play the game.

Thus, the actual game that individuals will play within the plea-bargaining institution depends on the motivations or preferences of the players. If one assumes that individuals have self-interested motivations, the game form of Table 1 becomes the classic prisoner's dilemma portrayed in Table 2.[35]

35 The prisoner's dilemma portrayed in Table 2 presents a payoff structure that might be

Table 2
The Plea Bargaining Institution Played by Pure Egoists:
The Prisoner's Dilemma

ROW/COLUMN	Cooperate	Defect
Cooperate	2,2	0,3
Defect	3,0	1,1

One derives Table 2 from Table 1 by substituting in each cell of Table 2 the agents' evaluations of the outcome of the corresponding cell in Table 1. The first entry in each cell represents Row's *self-interested* evaluation of the outcome of that cell; the second entry represents Column's *self-interested* evaluation of that outcome. So, for example, each agent ranks the outcome (6 months, 6 months) from the strategy pair (cooperate, cooperate) as second best of the four possible outcomes. In the unique Nash-equilibrium of this game — and an equilibrium in dominant strategies — each player chooses to defect, i.e., to confess. The players therefore each receive a sentence of three years, while each player would prefer the outcome in which both cooperate — i.e., remain silent — in which each receives a sentence of six months.

A completely different game results from the institutional structure if Row and Column are purely altruistically motivated — i.e., they care only about the payoff to the other party.[36] The new strategic situation, presented in Table 3 below, derives from the institutional structure of Table 1 in the same manner as the classic prisoner's dilemma: one evaluates the outcomes in

---

reached from a variety of different outcome structures, depending on the motivations of the players. Phrased differently, the prisoner's dilemma game is independent of the motivations of the players. The payoffs in each cell represent each player's ranking of the particular outcome (relative to the other outcomes); they do not depend on the agent's evaluative criteria — whether self-interest or altruism. Thus, altruists as well as egoists might confront the dilemma posed by the game.

36 Note that the two games presented in Tables 2 and 3 do not exhaust the possible games that would result from the institutional structure of Table 1. For instance, one agent, say Row, might have altruistic preferences, while Column has self-interested preferences. This assignment of motivations would lead to a different strategic structure.

each cell of Table 1 according to the purely altruistic preferences of the agents. Consider, for example, the outcomes from the strategy pair (cooperate, defect) that results in the outcome of five years for Row and probation for Column. If Row has self-interested preferences, as in Table 2, she will rank this outcome last, but if she has altruistic preferences, she will rank this outcome first.

Table 3
The Plea-Bargaining Institution Played by Pure Altruists: The Prisoner's Delight

ROW/COLUMN	Cooperate	Defect
Cooperate	2,2	3,0
Defect	0,3	1,1

In Table 3, each player has a dominant strategy: cooperate. As a consequence, the unique equilibrium is the "socially" desired one. Notice that the equilibrium of the prisoner's delight in Table 3 produces the outcome that is best not only from the point of view of pure altruists, but also from the point of view of pure egoists; the plea-bargaining institution performs better when populated by altruists. More importantly, from the perspective of institutional design, whether an institution gives rise to the strategic situation in Table 2 or to the strategic situation in Table 3 depends on the motivations of the population within the institution. After all, an institution determines objective outcomes as a function of the actions of the players, while the strategic structure of the situation is determined by the players' subjective evaluation of these outcomes.[37]

The dependence of institutional outcomes on the motivations of actors within the institution is a widespread phenomenon that appears in more complex models than the simple prisoner's dilemma. Consider Glazer's

37 Amartya Sen, *Behavior and the Concept of Preference*, 40 Economica 241 (1973), suggested that the prisoner's dilemma would be resolved if each player sought to maximize the preferences of the other player. The formulation in the text puts Sen's point in the proper formal context.

and Rubinstein's model of arbitration.[38] A panel of expert judges must issue an opinion concerning some case. Each judge might have a public motive, a private motive, or a mixture of the two motives in deciding the case. If she has a pure public motive, she cares only about the maximization of the probability of correctly deciding the case. If she has a purely private motive, she cares only about the probability that her vote coincides with the selection of the panel.

If judges have purely private motives, then there is no mechanism to implement the first-best outcome. If judges have purely public motives, then there are multiple equilibria, at least one of which is non-optimal. But if each judge has mixed motives and cares about having her opinion accepted, then the unique equilibrium implements the first-best outcome.

These examples strongly suggest that the assumption that incentive-based institutions will perform well even when inhabited by virtuous agents is false. This defense of the use of *homo economicus* as the behavioral theory for questions of constitutional design also fails.

F. An Example: Electoral Institutions

The discussion thus far has been abstract. Brief examination of a concrete issue in constitutional design may further the argument. I thus propose to consider the design of electoral institutions. These institutions are fundamental to any polity, as they determine the membership in all elected offices. Moreover, the analysis of voting rules carries over to some issues in the design of legislative institutions, as these too operate through the votes of the elected representatives.

The design of electoral institutions requires the joint specification of several features. First, one must determine the size of the legislature (or of the chamber of a multi-cameral legislature). Second, one must determine the districts from which members will be elected. Third, one must determine how many representatives each district may elect. Finally, one must determine the voting rule used in each district.

Consider first voting rules. Notice that the assumption of *homo economicus* is both too strong and too weak. It is too strong because one only has to specify that agents seek to maximize their preferences in order to analyze the voting rules; one need not assume self-interested preferences. The assumption of self-interested preferences is too weak, because, in general, it does not identify a unique, Nash equilibrium to a voting game. Indeed,

---

38 Jacob Glazer & Ariel Rubenstein, *Motives and Implementation: On the Design of Mechanisms to Elicit Opinions*, 79 J. Econ. Theory 157 (1998).

in general, there will be an infinite number of equilibria with virtually any outcome possible. So the assumption of strong rationality here does not allow prediction, unless the analyst further restricts agent behavior.

Self-interested rationality, however, presents further problems. It does not, for example, provide a convincing account of an individual's decision to vote. In large electorates, the expected benefit of casting a vote is very small for any given voter because the probability that the individual will be decisive in the election is negligible. People nonetheless vote; in some electorates, the vast majority of eligible voters in fact cast a ballot.

Self-interested rationality will also provide indeterminate answers to the other questions concerning electoral institutions. The decennial reapportionment process in the United States illustrates the difficulty in articulating standards to guide the drawing of districts in a non-partisan manner.

Similarly, self-interested rationality does not determine what the content of politics will be within a given jurisdiction. When conflict within a polity is restricted to a single dimension, the government is more stable.

Despite the failures of self-interested rationality to guide the choice of electoral institutions, the analyst has an empirical basis on which to make design decisions. Duverger's Law identifies a relationship between stability in government and the number of parties. The fewer the parties, the more stable the government. Moreover, first-past-the-post, single-member districts are known to lead generally to two-party systems.

CONCLUDING REMARKS

In this essay, I have examined some of the foundations of constitutional political economy. I have argued that these foundations are shaky in two important respects. First, each of the varied defenses of the use of *homo economicus* as the behavioral model fails. Second, the normative criteria for the assessment of institutional arrangements require significant development.

Consider the choice of behavioral theory. Self-interested behavior is not the worst behavior, so that its use cannot be defended on grounds that extreme risk-aversion is the appropriate response to uncertainty over the motivations of agents. Further, the fact that some public officials will act in a self-interested fashion does not justify design on the assumption that public officials will *all* have this type of motivation. Self-interested behavior does not necessarily drive out other behaviors or motivations. Finally, incentive-based institutions such as markets do not always perform reasonably when the agents within those institutions have non-self-interested

motivations. Incentive-based institutions are thus not necessarily robust to variation in the motivations of agents. These conclusions suggest that constitutional political design must investigate a wider range of motivations than has typically been considered.

Difficulties in the identification of evaluative criteria have several sources. First, constitutional design presents several problems, rather than a single problem. The argument here has distinguished two such problems: an ideal problem and a design problem. Each problem makes different assumptions about the knowledge and motivations of the *designer.* In the ideal problem, the designer acts under extreme uncertainty, but this uncertainty makes the prediction of the consequences of different constitutional arrangements difficult to predict. These difficulties are certainly problematic for a designer who seeks to maximize social welfare, but they also present problems for consent theories when consent relies on preference. In the design problem, the extent of uncertainty is less, but the prediction problems remain massive. Moreover, consent theories confront a second difficulty: as agents know their own positions at the time of design, the normative appeal of consent is weakened.

These questions about the foundations of constitutional political economy, however, should not blind us either to the insights it has already provided into the functioning of political institutions or to its promise for further understanding of political organization.