# Models-As-Fables: An Alternative to the Standard Rationale for Using Formal Models in Political Science

James Johnson

Political scientists invoke the standard rationale to justify making and using formal models. It goes like this: (1) we rely on formal models to generate predictions, (2) we treat these predictions as empirical hypotheses, and (3) we seek to test these hypotheses against evidence derived from the "real world." I show that this interpretation of formal models as directly empirical is inadequate just insofar as it fails to capture the way we actually use them. I then offer an alternative rationale for making and using formal models. Specifically, I argue that we use models, like we use fables, for conceptual purposes.

> Mathematical models in social science are like fables … that we read to get insights into the social world in which we live.
> —Roger Myerson

> The word "model" sounds more scientific than "fable" … although I do not see much difference between them.
> —Ariel Rubinstein

Many political scientists will regard my epigrams as off-the-cuff comments, to be humored, perhaps, but not taken seriously. That dismissive judgment is a mistake. I defend the view that in making and deploying formal models we construct fables and we do so in order to convey relatively simple lessons, morals, or "truths." This in no way subverts the scientific aspirations that many in the discipline embrace. Instead, my arguments help clarify those aspirations.

I claim no originality for the suggestion that we think of models as fables. Prominent game theorists repeatedly draw this comparison, although without pursuing their remarks in a systematic way.[1] But their observations converge with more sustained arguments advanced by philosopher Nancy Cartwright (1999, 2010). As a point of departure, I believe we ought to take these accomplished economists and philosophers at their word when they insist that when we make formal models, we really are constructing fables.

In pursuing this line of inquiry, we will depart markedly from what I call "the standard rationale." This is a view that prevails in contemporary political science. It goes like this: (1) we rely on formal models to generate predictions, (2) we treat these predictions as empirical hypotheses, and

*James Johnson is Professor of Political Science and teaches political theory at the University of Rochester (jd.johnson@rochester.edu). His research traverses pragmatist political thought, democratic theory, philosophy of social science, and political economy.*

*This paper has had an extremely long gestation. In 2016 I presented an initial, half-baked version at the Department of Political Science. University of Geneva (April). I presented a revised version in 2017 on a panel at the Midwest Political Science Association meetings (April) and at a workshop: "What to Make of Highly Unrealistic Models?" at the University of Helsinki (October). In 2018 I presented re-revised versions at the Departments of Political Science, Trinity College, Dublin. (April) and University of California, San Diego (June) and then in the Blalock Lecture Series, at the ICPSR Summer Program in Quantitative Methods of Social Research (July). The folks who arranged all of those events know who they are and how grateful I am for the opportunities they afforded me. To everyone who criticized and questioned my views during one of these events, thank you.*

*Starting in the fall of 2018 the manuscript was desk rejected by editors at several journals. Those folks know who they are too. No hard feelings. Here at* Perspectives on Politics*, I want to thank Michael Bernhard for his willingness to take an intellectual risk, for his patience with my glacial revision pace, and for coordinating a rigorous review process. I also want to thank four referees for their frank, skeptical comments. In particular I want to thank Referee #3 for their deep, abiding disagreement. I really do hope you will argue back.*

doi:10.1017/S1537592720003473



© The Author(s), 2020. Published by Cambridge University Press on behalf of the American Political Science Association

https://doi.org/10.1017/S1537592720003473 Published online by Cambridge University Press

(3) we seek to test these hypotheses against evidence derived from the "real world." Models, according to the standard rationale, are valuable for directly empirical purposes. The considerable distance between the standard rationale and the models-as-fables view poses no difficulty, I will claim, just insofar as the former, on examination, is thoroughly unpersuasive.

I offer two opening caveats. First, I deal exclusively with "formal" models (e.g., game theory, social choice). I deal only glancingly with "statistical," "causal," or "empirical" models of various sorts or with the diverse relationships between these and models of the formal sort.[2] In what follows, "models" is shorthand for formal models.

Second, political scientists tend to assume that we properly assess the social sciences primarily, perhaps exclusively, in terms of empirical performance. This shared premise is indefensibly simplistic.[3] For some time, philosophers of science have argued, quite plausibly, that when we judge the practice and progress of scientific inquiry we appeal to "multidimensional" standards. This means that our overall judgements will incorporate appraisals of the ways scientists improve or perfect instruments and of the ways they address conceptual problems, in addition to assessments of sheer empirical performance (e.g., Kitcher 1993, 90–126; Laudan 1977; 1996). This recognition is crucial to my argument because I claim that we use models as tools of conceptual exploration. I take this multidimensional view as established and do not defend it in what follows.

With those caveats in mind, I argue that the standard rationale is unpersuasive on simple empirical grounds. Many of the most important and esteemed models in political science make no predictions whatsoever. The standard rationale persists less because it accurately captures actual practice than because we lack an alternative. If there is value in formal modeling—and nothing I say here suggests otherwise—then we need a more perspicuous view of what that value consists in. I aim to offer such a clearer view.

I want to be clear about the nature of my argument. I think it is crucial for a range of intellectual and practical reasons to dislodge the standard rationale. And I am confident that the alternative models-as-fables view affords a persuasive interpretation of how we actually use formal models. Some readers may find both claims compelling. That said, I hardly expect those committed to the standard rationale to welcome my argument. Nor do I expect all who may accept my criticisms of the standard rational to embrace the alternative I advocate. I hope only that readers who fall into either of those camps will accept my argument for what it is—not just a challenge but a provocation. I aim to provoke such readers to defend the standard rationale in a compelling way or to advance yet another, more persuasive alternative to it.

The remainder of the paper consists in seven sections. I begin by offering reasons to think that what I depict as "the standard rationale" is indeed standard. I then illustrate why the standard rationale is implausible by examining at some length Positive Political Theory (PPT) as it emerged at the University of Rochester over three-plus decades. I show that while advocates of that research tradition embrace the standard rationale and while their central intellectual contribution to political science emerges from a sophisticated use of formal models, they did not use models in the way the standard rationale suggests. In other words, there is a marked discrepancy between the tales advocates of PPT tell about their models and the way they, in fact, have produced and used them. I believe that, rather than struggling to bring formal models into conformity with the standard rationale, political scientists ought to reject the prediction and testing interpretation of the modeling enterprise.

After demonstrating the inadequacy of the standard rationale, I defend a view of models-as-fables as an alternative interpretation for how political scientists use formal models. This involves establishing commonalities between models and fables and specifying how we use models to show what abstract concepts (e.g., rationality, institution, etc.) mean in relatively concrete circumstances. In the conclusion I underscore the force of the reconstruction that I propose.

## The Standard Rationale

The excessive disciplinary preoccupation with empirical performance sustains "the standard rationale" for constructing and solving formal models. As already noted, on this view, formal models are used for more or less directly empirical purposes. Rebecca Morton offers arguably the most sustained defense of this view. She *defines* "model" (whether formal or informal) in terms of the task of generating predictions about the "real world" (Morton 1999, 61; Morton 2009). But she is neither the only, nor the last to do so (Dowding 2016, 79-88; Dowding and Miller 2019).

Perhaps more tellingly, modelers themselves articulate the standard rationale when their work departs from it. Consider this pair of apt examples:

> Much of modern social science research centers on an idealized version of the scientific method in which a model is built on first principles, predictions are derived from this model, and these predictions are then subjected to empirical verification. (Patty and Penn 2014, 10,78,121)

> Normally the ideal paper for the mathematical modeling crowd is a well-specified game that reaches some equilibrium outcome, which is then instantiated and tested in an appropriate statistical model. (de Marchi 2005, 2)

In both instances, the authors clearly articulate the standard rationale in order to highlight that they themselves plan in one or another way to depart from it. Convention here appears most succinctly in the breach.



https://doi.org/10.1017/S1537592720003473 Published online by Cambridge University Press

The standard rationale informs research practice in political science. Consider three examples. First, it informs the current preoccupation with descriptive and causal inference. This is true both when advocates of causal inference take it as grounds for dismissing formal models (e.g., King, Keohane, and Verba 1994, 105-7) and when they seek to show how formal modeling is fully compatible with the turn to inference (Monroe et al. 2015). Second, the standard rationale informs many defenses of experimental methods as especially useful for (among other things) testing predictions generated by formal models (Aldrich and Lupia 2011, 90; McDermott 2002, 31; Morton and Williams 2008, 349–50; Woon 2012, 54-6). Finally, the prediction-and-testing view of formal models embodied in the standard rationale affords arguably the most common interpretation of the influential training program on Empirical Implications of Theoretical Models (EITM). Advocates consistently depict the initiative in terms of building formal models and "evaluating" them empirically.[4]

These examples should suffice to warrant my depiction of the standard rationale for using formal models as indeed *standard.* Not every political scientist subscribes to this rationale (e.g., Clarke and Primo 2012). And those who endorse it often do not do so explicitly, exclusively, or consistently. However, when discussing formal models in political science, both advocates and critics of the enterprise invoke criteria of empirical performance generally and predictive capacity specifically to justify their assessments.

## The Predicament of "Positive Political Theory"

The standard rationale encounters obvious and widespread difficulties. Political scientists commonly use models in ways that depart from the standard rationale even as they persist in invoking it to justify their work. To illustrate this, I examine the research tradition known as "Positive Political Theory" (PPT) as it emerged from the University of Rochester in the 1970s. I focus in particular on seminal work revolving around models of disequilibrium and institutions in politics.

Advocates of PPT are inveterate tellers of tales. This occurs most obviously in textbooks, whether elementary (Riker and Ordeshook 1973; Ordeshook 1992; Shepsle 2010) or more advanced (Austen-Smith and Banks 1999, 2005; McCarty and Meirowitz 2007; Ordeshook 1986). But it also occurs in review essays (Austen-Smith and Banks 1998; Krehbiel 1988; Schofield 2001; Shepsle 2006; Diermeier 2014a,b) and historical reflections (e.g., Amadae and Bueno de Mesquita 1999; Bueno de Mesquita and Shepsle 2001; Riker 1977; 1990). Unsurprisingly, not all these tales are consistent with one another. Indeed, contending tales reveal serious disagreements among advocates of PPT, albeit less about their shared commitment to the standard rationale than about how best to fulfill that commitment.

None of this, in itself, is a criticism. Such telling of tales is central to the elaboration and defense of any research tradition (Laudan 1996, 142-53). Nor is there any embarrassment to the fact that members of a given research tradition disagree among themselves (MacIntyre 1977, 461). My own aim is to present a competing tale, one that drives a wedge between PPT models and the standard rationale. It nevertheless *is* a criticism to point out when and how the tales advocates of a research tradition tell depart from the actual intellectual trajectory and accomplishments of their joint enterprise. That is the sort of criticism to which advocates of PPT lay themselves open. They repeatedly portray their enterprise as guided by allegiance to the standard rationale. But prediction and testing played at most a marginal role in the research that represents their single most important intellectual contribution to political science.

The following characterization of PPT affords one auspicious point of departure:

> The goal of positive political theorists is to build models that predict how individuals' self-oriented actions combine to yield collective outcomes… . Using game theory and formal models, positive political theorists strive to determine whether … complex, strategic political interactions have predictable, law-like outcomes that exhibit stability. Stable outcomes, referred to as equilibria, signify that agents' actions combine in such a way that, given the collective social outcome of agents' self-oriented actions, no individual could achieve a greater (expected) payoff if he had unilaterally selected an alternative course of action. Equilibria are significant to positive political theorists because they indicate that the political processes under investigation result in predictable, stable social outcomes that best serve individuals' constituent interests, given the constraints imposed by the situation. The sequence of strategic choices that form an equilibrium and that imply specific outcome events constitutes the core of a predictive science of politics. (Amadae and Bueno de Mesquita 1999, 270-1)

And, lest the reader misconstrue their claim, the authors insist that as an historical matter the task of creating formal models to generate empirical predictions is just what differentiated PPT from the allegedly benighted state of mid-century political inquiry. Indeed, they invoke the standard of predictive performance to punctuate their sketch of the characteristics distinguishing PPT from alternative modes of inquiry (Amadae and Bueno de Mesquita 1999, 287-9).

On this view, the enterprise of PPT obviously and necessarily embodies the standard rationale. Yet the way PPT advanced bears only a tenuous relation to this tale. Amadae and Bueno de Mesquita suggest that in its formative stages the research tradition explored first spatial, then game theoretic models of the dimensionality of political decision-making and incorporated that concern into efforts to model the crucial role of institutions in politics. This, they argue



https://doi.org/10.1017/S1537592720003473 Published online by Cambridge University Press

emerged as "a core area of inquiry" among advocates of PPT (Amadae and Bueno de Mesquita 1999, 283).

The problem with this reconstruction is simple: *none* of the crucial advances in the important line of PPT research into models of disequilibrium and institutions relied in any substantial way on the pattern of prediction and empirical testing summarized in the standard rationale.

## Retelling Rochester

The formal analysis of political disequilibrium and institutions produced by political scientists affiliated with the Rochester department is an immensely important intellectual legacy. In practice, the insight and understanding we derive from formal models is extremely valuable. It simply turns out that such insight does not derive from prediction and empirical testing as the standard rationale stipulates. Once we establish a clearer view of that, many objections to the Rochester analysis of institutions become moot. Moreover, once we do so it is possible to advance a more persuasive rationale for why political scientists use models more generally.

There are four crucial turning points in the Rochester tale: first, the models of generalized instability of majority rule; second, the formal analysis of institutional equilibria; third, the "Riker Objection" arguing that institutional equilibria are themselves susceptible to generic instability; and, finally, the game theoretic analysis of equilibrium institutions. Each step amounts to an intervention in an ongoing argument over what counts as rationality in politics. None involved enacting the standard rationale. In this section I defend that assertion by sketching each step in this extended intellectual enterprise and what it involved. I then examine how those committed to PPT have responded by trying to rescue the standard rationale.

### *Generalized Instability of Majority Rule*

The foundational impetus to PPT analyses of institutions emerges from models that examine political interactions in environments that are institutionally sparse. These initial models capture central features of collective choice procedures such as majority rule.

It will help to provide a bit of conceptual background. An aggregation rule (say, majority rule) provides a method of mapping from individual preferences or judgements to a collective decision. Social choice theorists refer to the set of maximal or socially best outcomes for a given aggregation rule as the *core.* In a single dimension setting, the majority rule core consists of a Condorcet winner—an alternative that defeats each other available option in a series of pairwise comparisons. In a multi-dimensional context, however, social choice theorists have established formally that the conditions for the existence of a core are exceedingly restrictive. Indeed, in many instances the core is empty. For present purposes that is our point of departure.

McKelvey (1976, 1979) and Schofield (1978) establish that in multi-dimensional settings when the core is empty, when voters have Euclidean preferences, and when they vote sincerely, preference aggregation by simple majority rule will not generate an equilibrium so long as voters retain agenda-setting power. This is a characterization of majority preference, not a representation of political behavior. Under the conditions noted, from any status quo, there exists a finite path of majority decisions leading to any point in the policy space. In other words, "when social preference breaks down, in the sense of not admitting a core, it breaks down completely in that one can get from any one alternative in the space to any other via the social preference relation" (Austen-Smith and Banks 1999, 178). This is what it means to talk of the generalized instability of majority rule.

McKelvey and Schofield often are depicted as having advanced "chaos theorems" that describe a fundamental pathology of democratic politics. Indeed, some advocates of the standard rationale depict the McKelvey–Schofield results as "disequilibrium predictions" and, hence, as at least in principle susceptible to empirical testing (Morton 1999, 102, 182–83, 208, 281). That interpretation, however, is as tendentious as it is common.[5] To see why, consider a second irony. Not only was the PPT work on institutions prompted by collective choice models that are institutionally austere, those models actually generate no predictions whatsoever. On this interpretation, it is "naïve" to claim that McKelvey and Schofield "predict political chaos." Such a view "attributes a positive prediction to a model that does not make any" (McCarty and Meirowitz 2007, 79-80).[6] There is no embarrassment in that. Recognizing that the McKelvey–Schofield models make no empirical predictions poses a problem only if one is committed to the standard rationale.

### *Institutional Equilibrium*

Working within the framework of collective choice theory, Shepsle (1979) initiated a stream of research on what he called structure-induced equilibrium. If systemic instability haunts the institutionally austere McKelvey–Schofield environments, Shepsle shows how in legislative settings, but not only there, various institutional rules (e.g., jurisdictional assignments and amendment procedures) operate to constrain instability. By incorporating institutional features of collective choice, models can generate structure-induced equilibria (SIE) or what can be thought of as institutional equilibria. Yet while Shepsle relies throughout on casual empiricism and mathematical examples to motivate his argument, his model makes no predictions. Rather, he establishes a set of theoretical/conceptual results, offering (1) a characterization of how various institutional rules can generate SIE and (2) a discussion of the conditions under which SIE will exist. The departure from the standard



https://doi.org/10.1017/S1537592720003473 Published online by Cambridge University Press

rationale is clear. What is more, the departure is not confined solely to Shepsle's initial work.

Shepsle was among the very first in a flurry of formal models of political (mostly legislative) institutions. Krehbiel (1988) offers a detailed and lucid survey of this work through roughly its first decade. His account is useful precisely insofar as it reveals that the research did not develop by conforming in any noticeable way to the standard rationale. Krehbiel (1988, 260) states at the outset that assessment of spatial models of legislative choice rests on the extent to which the models survive "empirical scrutiny." But he in fact has nearly nothing to say about the empirical performance of the models he surveys. Indeed, he punctuates his review with what advocates of the standard rationale must consider a damning assessment. After a decade of elaboration, spatial models of legislative politics, Krehbiel (1988, 313) assured us, were "approaching testability" even if they had yet to attain that aim. I return to this observation later. Here it is important to note, once again, that this conclusion is embarrassing only if one subscribes to the standard rationale.

### *The Riker Objection*

Shepsle and the work on spatial models of legislative institutions elicited what has been called "the Riker objection" (Shepsle 2006). In a well-known commentary Riker (1980) leveled a fundamental challenge to the notion that "institutions matter." He argued that institutional equilibria are susceptible to precisely the dynamic that generates instability when institutions are absent. On his view, institutional equilibria are the arbitrary result of interactions among individuals. As such they can and, crucially on his view, *will* be altered if some subset of those individuals find themselves relatively disadvantaged by extant institutions in an ongoing way (Riker 1980, 445). Losers, in other words, have good reason to try to manipulate or alter institutional rules.

There is, on Riker's account, no reason to think that the disequilibrium McKelvey and Schofield reveal is confined to decisions over policy. Conversely, there is good reason to see choice of institutions simply as a special case of policy choice. According to Riker: "If institutions are congealed tastes and if tastes lack equilibria, then so also do institutions, except for short-run events" (Riker 1980, 445). This pointed criticism took direct aim at the explanatory claims made by Shepsle and others. But note that this is a *conceptual* claim. Like Shepsle, Riker is probing the character of unobservable causal mechanisms and the conditions under which they operate. Yet his complaint does not rely for its force on anyone, whether himself or those he is criticizing, having derived equilibrium predictions and tested them empirically.

### *Equilibrium Institutions*

In a remarkable pair of papers Calvert (1995a, b) offers a powerful response to Riker's objection. He formulates a general game theoretic argument in which institutions emerge as sets of rules sustained as equilibria in an underlying repeated game (Calvert 1995a, 228–33; 1995b). In other words, Calvert offers a characterization of "equilibrium institutions" (Shepsle 1986) to account for why institutions persist in the face of the concerted discontent of those who consistently lose under existing rules.

Calvert models the basic underlying game as an indefinitely repeated, N-person prisoner's dilemma in which players are randomly matched for pairwise interactions. He elaborates increasingly complex versions of this model in each of which he identifies threshold values for the discount parameters of players, the size of N and, eventually, the costs of communication that sustain a variety of cooperative equilibrium outcomes, both non-institutional and institutional. These equilibria all involve some more or less elaborate variation on conditional cooperation. They range from what we might call specific reciprocity in which mutual cooperation remains non-institutional, through the sorts of generalized reciprocity characteristic of informal social norms, and on through increasingly formalized and centralized institutions that incorporate first multilateral communication, then centralized communication, and finally centralized communication coupled with protections against higher order malfeasance.

These models are neither empirical nor properly explanatory. Calvert notes as much. His models do not account for the emergence of *any* particular social or political institution. They are an analytical exercise aimed at producing an existence result, one that establishes the possibility of equilibrium institutions, specifying in the process the precise conditions under which such equilibria might emerge. The difficulty, of course, is that as in nearly any repeated game where players discount the future, each of the series of games he examines generates multiple equilibria. This unavoidable coordination problem sets the basic explanatory task for institutional theories. Which, if any, of the available equilibrium institutions actually will emerge from social and political interaction? How might we best explain the way institutions are generated? Under what conditions will such institutions persist and operate over time? These questions are central to the explanatory ambitions of PPT theories of institutions (Calvert 1995a, 228). They are questions that Calvert explicitly does not address and regarding the possible answers to which he remains agnostic (Calvert 1995b, 70, 80–82).[7]

Calvert, on this interpretation, persuasively answers Riker. He answers one conceptual claim with another. In the process, he opens a range of crucial explanatory questions. But for present purposes it is most important to note that prediction and testing play no more a role in this advance in the PPT research tradition than they did at earlier turning points.

Put bluntly, the standard rationale played no discernable role in the fundamental PPT arguments over political



https://doi.org/10.1017/S1537592720003473 Published online by Cambridge University Press

disequilibrium and institutions. Indeed, various parties to the argument relied on models for *conceptual* purposes. This in no way diminishes their collective intellectual accomplishment, which affords foundational resources for how we think about political institutions. It is more plausible to claim that it is commitment to the standard rationale that suffers, or ought to.

Unsurprisingly, advocates of PPT do not derive anything like this lesson from the intellectual history I have reconstructed. However, I am hardly the first to have noted a discrepancy between the allegiance advocates of PPT proclaimed to the standard rationale and the actual practice members of the research tradition engaged in as they explored models of disequilibrium and institutions in politics. Contemporary advocates of PPT seek in different ways to rehabilitate the research tradition to bring it into line with the standard rationale. I here consider one such effort.

### *Institutionalism as Methodology*

In a provocative revisionist account Diermeier and Krehbiel (2003) identify the discrepancy I have sketched with special clarity and considerable dissatisfaction. In response, they reassert the standard rationale and chastise fellow proponents of PPT for departing from it in practice. They insist it is a mistake to interpret "institutionalism" as a set of substantive insights about politics as McKelvey, Schofield, Shepsle, Riker, and Calvert all do. They suggest instead that we treat institutionalism as a "methodology" that distills the standard rationale.[8] They advance this interpretation of "institutionalism" in hopes of reconciling the recalcitrant intellectual history of PPT with steadfast commitment to the standard rationale.

Diermeier and Krehbiel attribute the discrepancy between the actual trajectory of PPT and the precepts articulated in the standard rationale to a basic methodological mis-judgement. Too much of the work on political disequilibrium and institutions, in their estimation, depended on models of direct preference aggregation. They distance themselves from such collective choice models and ground their "institutions as methodology" program in methods of non-cooperative game theory.[9] The latter afford the appropriate tools for generating empirically testable predictions.

Perhaps the best way to see how this effort to redeem the standard rationale fares is to revisit Krehbiel's assessment, discussed earlier, of the initial decade or so of research on institutional equilibrium models. You will recall that he concludes his survey by suggesting that, at that time, formal models of legislative politics were "approaching testability." It is only slightly uncharitable to suggest that, after a further decade and a half, that promissory note remained unredeemed. Krehbiel, Meierowitz, and Woon (2005) recognized this shortcoming and aimed to remedy it. They sought to "extract falsifiable predictions" from competing spatial models of legislative politics. The task proved daunting. Despite much determined, creative effort Krehbiel and his coauthors found it difficult to discriminate the "predictions" they sought to extract from competing formal models of legislative politics in the United States—either from baseline models or from one another. As a result, their sophisticated effort to "test" competing models proves, by their own admission, "inconclusive" (Krehbiel, Meierowitz, and Woon 2005). As they noted of the extant models, "their predictive abilities seem neither very impressive nor very different from one another" (Krehbiel, Meierowitz, and Woon 2005, 252). Yet "testability" minimally presupposes that competing models generate robust, distinct predictions. Despite the best efforts of Krehbiel and his co-authors, it is unclear that that condition has been met. It is only commitment to the standard rationale, however, that might prompt us to interpret this as a shortcoming.

A plausible alternative interpretation of this state of affairs is that the models in question simply do not generate empirical predictions in any meaningful sense. On this view the difficulty does not reside, as Krehbiel and his co-authors suggest, merely with technical difficulties surrounding models, methods, and data. It arguably resides instead with an underlying category mistake: assuming in the first place that we should interpret formal models in terms of the standard rationale.

## Theories, Models, Interpretations

Political scientists typically take the standard rationale to be self-evident. That is a remarkable presumption given the actual intellectual trajectory of PPT. In the remainder of the paper I offer an alternative interpretation, one that sets aside the preoccupation with prediction and testing that constitutes the standard rationale and instead treats models as fables. The first step here is to explain why we should relinquish the assumption that models are empirical in any direct way.

Advocates of the standard rationale *assume* that formal models are empirically testable. They thereby take as settled matters that actually remain thoroughly contestable. To understand why this is so we must differentiate between theories and models. Having done that, we must recognize the range of possible interpretations on offer of models and why we use them. Distinguishing theories and models allows us to ask how the two are related and inquire as to whether models are empirical in the first place.

Among philosophers of science of various stripes something of a consensus exists that whatever else we might want to capture with the idea, a *theory* consists minimally in a set of claims about the world, meaning a set of claims with substantive content. So understood, a theory can be assessed empirically and found to be supported or not by appropriate evidence (Hacking 1983, 175; Hausman



https://doi.org/10.1017/S1537592720003473 Published online by Cambridge University Press

1992a, 27; Rosenberg 2005, 69). I take this shared understanding as given.

What then is a *model*? This question is poorly posed in at least two ways. First, it takes little reflection to grasp that "there are lots of kinds of models serving lots of different purposes" (Cartwright 1999, 180). It therefore seems implausible to imagine that there exists a single set of definitional criteria which any "model" must meet. Second, asking what a model *is* diverts attention from what for present purposes are more germane questions: What are models for? For what purposes do we use models in the course of inquiry?

Political scientists primarily use two broad sorts of formal model. These can be differentiated in terms of their "methodological approaches" (Austen-Smith and Banks 1998, 263). The first consists in models of "direct preference aggregation," often relying on "methods" of social choice theory. The second focuses on the vicissitudes of strategic interaction and relies on "the methodology of game theory" (Austen-Smith and Banks 1998, 260-61). So, while political scientists commonly speak of game *theory* or social choice *theory* it seems more accurate to dispense with the honorific and see these undertakings as simply loose "families of models" (Schelling 1978; Austen Smith and Banks 1998).

Advocates of PPT concede that models of direct preference aggregation perform poorly in terms of generating predictions that might be treated as empirical claims about the world (Diermeier and Krehbiel 2003; Austen-Smith and Banks 1998). It therefore hardly seems apt to inquire about the substantive or empirical content of this family of models. I focus instead on game theoretic models since that is where proponents of the standard rationale seem to have cast their lot. That said, it is easy enough to see that game *theory* is not one, at least in the minimal sense noted earlier. This is so because it lacks the requisite substantive content. As one practitioner bluntly states: "Game theory is a method of analyzing strategic interactions, not a substantive theory of some particular class of interactions."[10] In that sense it "has only a little more substantive content than does statistical theory" (Cox 1999, 158). So, while this methodology is, I think, extremely useful, it is not in itself a theory of politics or of anything else.

There is the conundrum. On the one hand, we have a generic idea that a "theory" involves substantive claims about the world. On the other hand, the most likely candidate for a formal or positive "theory" of politics is a methodology - a family of models, nearly devoid of substantive content. Conversely, a theory surely is not simply a list of hypotheses either. Yet drawing some relatively sharp contrast between models and theories is useful. It allows us to explore how theories and models are related and the purposes for which we might use the latter.

There are at least three broad views on these matters. Two of these—the syntactic and semantic interpretations—should be familiar (Clarke and Primo 2012, 62-79). I refer to a third interpretation as the predicate view. It stands out sharply from the more familiar interpretations in two respects. It sustains a distinction between theories and models. And it views the latter not as directly empirical but as instruments of conceptual exploration.

The standard rationale is informed by a what philosophers call the *syntactic* interpretation of theories. On this view, a model is comprised of statements, typically in mathematical language, from which we derive and test predictions to determine if they are confirmed empirically. In its classic form this view has little to say about models and treats them as more or less incidental to scientific inquiry (Contessa 2011, 121; Frigg and Hartman 2012, 23-7). The syntactic view therefore is not terribly persuasive. First, it accords little weight to what positive political theorists actually do, namely make and analyze models. Second, the difficulties I sketched with PPT highlight a further problem—many canonical models do not generate any predictions at all. It may, of course, be possible to rescue the syntactic view from such inconvenient observations, and in the process, help sustain the standard rationale. But that task clearly faces substantial difficulties. Minimally, they invite us to explore rival ways to interpret the point of building models and using them.

Those critical of the syntactic interpretation often adopt instead a *semantic* view models. They recommend that we think of models as objects, specifically as maps, which, because they are partial, purpose-relative, and not susceptible of being true nor false, are best understood as useful or not for some task or other (Giere et al. 2005, 21-7; Clarke and Primo 2012, 63–72). The semantic view is compelling not the least because it underwrites a pluralist understanding of models and their uses. For present purposes, however, it raises two sorts of difficulties.

First, regarding the task of distinguishing models and theories the semantic view is nearly the converse of the syntactic interpretation. Whereas the syntactic view treats models as incidental to scientific inquiry, the semantic view reduces the idea of a theory to simply a collection of models augmented by speculation about whether and how the model "fits" the world (Giere et al. 2005, 21-7; Clarke and Primo 2012, 74). Political scientists who use game theory may find this feature of the semantic view attractive. Importantly though, it comes at the price of relinquishing the standard rationale insofar as it treats models as incapable of being true or false. It thus thoroughly deflates the imperative to "test" formal models. It also raises the question of how to square the notion that a theory just is a collection of models with the rough consensus noted earlier that a theory involves substantive claims about the world.

Second, the semantic interpretation shifts the focus of attention from such questions as "What is a model?" to such questions as "How do we use models?" Those who embrace it generally answer such questions by suggesting



https://doi.org/10.1017/S1537592720003473 Published online by Cambridge University Press

that in deploying models we exploit some facet or other of similarity between the model and the "real world" in order to *represent* the latter (Giere 2004). Thus, while Clarke and Primo (2012, 82-93) endorse a pluralist view of how models factor in political inquiry, they retain a role—albeit a reduced one—for prediction. Yet if models are not just objects, but a specific form of object (e.g., a map), it is difficult to see how they can generate predictions at all. We use maps to navigate. However they are not sets of assertions that, like predictions, can be true or false.

The problems with the semantic view do not stop there. Even if one wants to insist that maps offer grounds for predicting things, there is good reason to question the basic assumption of the semantic view, namely that models—at least formal models of the sort political scientists use—actually represent the "real world" in any straightforward sense (Gibbard and Varian 1978; Cartwright 1999; Norton 2004). This marks a significant point of difference between the semantic view of models and the view I turn to now.

A third interpretation, the *predicate* view, sees models as linguistic entities—as definitions or conceptions—that are distinct from, but can be incorporated into, theories (Hausman 1992a, 72–78). On this view, we do not use models to make statements or assertions about the world. Consequently, "It is a category mistake to ask whether they are true or attempt to test them" (Hausman 1992a, 78). Instead, on this view we use models primarily for conceptual purposes (Rubinstein 1991, 909; Elster 2007, 312). And our concepts of, say, rationality or institution or deterrence are neither true nor false except in some trivial sense.

For present purposes the predicate view has multiple virtues. First, it comports with the sort of "multidimensional" approach to assessing the practices and progress of sciences that I have taken as established from the outset. It allows us to understand what might be involved in the seemingly obscure idea of "conceptual" progress. Second, the predicate view grounds a plausible interpretation of the Rochester PPT work on the dimensionality of rational decision-making and of institutions in politics. This is important because, when assessed in terms of the standard rationale, that work would be considered thoroughly anomalous.

Whereas neither the syntactic nor the semantic interpretation clearly differentiates models and theories, the predicate view keeps the two distinct and so poses the question of how we use models in the process of articulating a theory. Hausman offers this example:

> The models that economists employ should be regarded as complex predicates or concepts or as definitions of such complex predicates or concepts. A model is of the same logical kind as a predicate, such as "is triangular," or as a definition of what it is to be triangular. One thus does not test a model, nor does one say that a model is true (except in the trivial way in which a definition is true). This is not, of course, to deny that one can use models to make claims about the world which may be tested and which may be true or false. A model of a perfectly competitive economy, for example, itself does nothing more than to define the theoretical term "a perfectly competitive economy." One can then use the term to make claims such as "Japan's economy is not perfectly competitive". (Hausman 1992b, 25)

The testable claims about the world Hausman mentions here constitute "theory" in social science. And while he focuses on how microeconomic models inform theory, his characterization applies equally well to the way political scientists deploy the families of models central to PPT.

To take one canonical example, Aldrich introduces a host of formal models in the second chapter of his *Why Parties? The Origin and Transformation of Political Parties in America* (2011 [1995]). He does so in order to define such concepts as "rationality," "disequilibrium," "institution," "collective action problem," and so forth. He uses none of these models to derive empirical predictions but instead relies on them to provide the conceptual scaffolding for his theory of how parties emerge and operate in democratic systems. He aims to bring empirical evidence to bear on that theory in quite familiar ways.

Aldrich, of course, occupies a central place in the lineage of Rochester PPT. But we need not restrict ourselves to that research tradition in searching for pertinent examples of how models inform theories. Consider Elinor Ostrom's *Governing the Commons* (1990). Here again, we find a handful of formal models in the opening chapters. Ostrom does not use them to derive empirical predictions but to articulate both the view she is arguing *against* and to illustrate the sort of political-economic possibilities her empirical inquiry reveals. She first introduces four simple models to depict variations on the received view—that in the face of commons problems individuals confront unrelenting tragedies, logics, and dilemmas that can be mitigated only by regulations promulgated by central authorities or by privatization and markets. She then uses yet another slightly more complex model to illustrate an alternative, namely the possibility that "a community of citizens" can, in a decentralized way, extricate itself from such predicaments without privatizing property rights or abdicating to a centralized authority. She aims to motivate us to see that such a community can create institutions that sustain "self-organizing and self-governing forms of collective action" (Ostrom 1990, 29, 25).

These models do not reappear in the remainder of the book. But they do inform the remarkable case-based theory Ostrom develops as she elaborates on the possibility that her fifth model captures. Overall, her inquiry "demonstrates a world of possibility rather than of necessity" (Ostrom 1998, 16). She aims to re-formulate the concept of a commons problem in order to "enhance the capabilities of those involved" to institutionalize their interactions so as to avoid "remorseless tragedies" (Ostrom 1990, 7). To that end, the enduring contribution of her theory has not been to deploy



https://doi.org/10.1017/S1537592720003473 Published online by Cambridge University Press

**Table 1**
**Conceptions of Models**

| Interpretation | Advocate? | Models as Linguistic Entities? | Can Models be True/False? | Model-Theory Distinction? |
|---|---|---|---|---|
| **Syntactic** | Satz-Ferejohn (1994); Dowding (2016) | Yes (Statements) | Yes | No |
| **Semantic** | Giere (2004); Clarke-Primo (2012) | No (Objects— 'Maps') | No | No |
| **Predicate** | Hausman (1992a,b); Johnson (2014, 2019) | Yes (Definition—Conception) | No | Yes |

the models as a means to formulate and test predictions but instead using them as a conceptual scaffolding on which to articulate principles of institutional design that afford practical guidance to communities confronting commons problems (Ostrom 1990, 88-102).

I will not pursue these examples further. In what follows I instead seek to show how the relationship of models to theory is both more complex than Hausman's sketch allows and more fruitful. Seeing models as fables renders the predicate view he endorses considerably more persuasive.

## How to Do Things with Fables

The syntactic and semantic interpretations of models are familiar and well articulated. By contrast, Hausman really only just sketches what I call the predicate view. Yet this view departs dramatically from its more familiar rivals.

Hausman is not alone among philosophers of science in rejecting both the syntactic and semantic views. Nancy Cartwright (1999, 184-5), for instance, rejects both. She criticizes both for embodying a "vending machine" conception of model-making—insert currency (e.g., assumptions, parameters, etc.), do the math, and *plunk* out drops the desired confection (e.g., a nice equilibrium) that can serve as the basis for "scientific prediction" (Cartwright 1999, 58-9,181-6). Cartwright thinks the vending machine conception comports poorly with the actual historical practices of the sciences. She complains that "the whole point" of vending machine approaches "is the elimination of creativity—or whim—in the use of theory to treat the world" where by "treat" she specifically is interested in intervening to change or transform social and political practices and institutions. Much of that creative activity centers around the way we use models to connect theory and the world. What seems clear, however, is that the standard rationale treats models as vending machines.

For Cartwright (1999, 179-80) "models mediate between theory and the world." This is a complex view, subject to considerable argument and contestation (e.g., Morgan and Morrison 1999). On the basis of this view Cartwright insists that models are essential if our theories are to be useful for making substantive claims about and, more importantly from her perspective, intervening in the world. Indeed, the domain of any theory extends just as far as certain of its associated models allow and no further (Cartwright 1999, 196,209-10,180-1). What this entails will become apparent presently. Here it is important to note that Cartwright attributes considerable autonomy to models relative to theories and raises the question of just how we use models to mediate between theory and the world.

Cartwright distinguishes between representative and interpretive models in terms of their respective uses. We construct the former "with the aid of theory to represent real arrangements and affairs that take place in the world." Think here of statistical models but also of both econometric and computational models. We construct the latter for conceptual purposes, more specifically, to render relatively "abstract" concepts—e.g., rationality, institutions, etc.—more "concrete" (Cartwright 1999, 3, 180-1). Think *here* of formal (e.g., game theoretic or social choice) models. When our theory incorporates an abstract *concept* such as "institution," we can rely on models of this sort to articulate some more concrete *conception* of "institution" as consisting of sets of rules for generating equilibria (Shepsle) or as itself an equilibrium in some underlying repeated game (Calvert).

We use interpretive models to show what some abstract concept "consists in" or "amounts to" in some more particular setting (Cartwright 1999, 41, 189). The model allows us to say, in a quite precise way, "in this setting and for these purposes *this* is what 'institution' consists in" and to consider the implications of doing so. Cartwright insists that the need to "fit out" abstract concepts in terms of more concrete conceptions in this way "marks an entirely commonplace feature of language" (Cartwright 1999, 40). Scientific inquiry on her account necessarily relies on "abstract concepts … whose relation to the world must be mediated by more concrete concepts."[11] I refer to the latter as conceptions.[12] These are "very specific in their



https://doi.org/10.1017/S1537592720003473 Published online by Cambridge University Press

form: the forms are given by the *interpretive* models of the theory." This lends precision to our theories but at the same time circumscribes their range of application. Abstract concepts are applicable as far as our interpretive models, and the conceptions they articulate, extend and no further (Cartwright 1999, 3).

Cartwright (1999, 35-48) suggests that the best way to grasp how we use interpretive models is to think of them as fables. Comprehensive analysis of fable as a literary genre sustains her case (Blackham 1985). Before we further explore the models-as-fables view, however, it will help to emphasize the sharp contrast between interpretive models, as Cartwright depicts them, and the more familiar semantic view that treats models in representational terms. Advocates of the semantic view suggest that we think of models-as-maps. If we accept that characterization, the contrast between Cartwright and the semantic view is stark. It operates on at least two dimensions, direction and domain.

First, when we make and use a map we move from concrete to abstract. We aim to ignore, amplify, eliminate, stylize, and so on various features of the "real world" to suit our purposes. To draft a *useful* map we necessarily distort the real world. Awareness of this predicament explains why advocates of the semantic view are preoccupied with the similarity or otherwise of our models to the world. But in this sense their view that we use models to represent the "real world" is the converse of Cartwright who, as noted, insists that when making and using interpretive models we move from abstract to concrete. We use models to interpret abstract concepts by articulating more concrete conceptions. We move, in other words, in nearly the opposite direction as the models-as-maps view that advocates of the semantic interpretation of models endorse.

Second, if we think of models-as-maps we are led to consider their "domain of usefulness" (Leamer 2012, 26-7,16-7,165). For instance, Clarke and Primo (2012) offer as examples three maps of Boston each useful for a different mode of travel—by mass transit, by automobile, or on foot—but next to useless for other modes. The examples illustrate the partial and purpose relative nature of maps, and thereby of models. Clark and Primo stress the various differences among the three maps. But in doing so they neglect what they have in common. They all are maps of Boston. In that sense they share a domain of usefulness defined by their edges. Designed to facilitate different modes of travel in Boston, they are of no use for guiding excursions in Framingham or Worcester, to say nothing of the wilds of the Berkshires. If models are maps, and so properly assessed in terms of their usefulness, they are useless outside a pretty obvious domain.

Here is the catch. As Leamer (2012 26-7,16-7) repeatedly points out, for the most part not only is "the domain of usefulness of our mathematical models … pretty much a mystery," it is a mystery about which practicing social scientists show scant interest. In this way the models-as-maps analogy common to the semantic view breaks down. We might take Leamer to be offering this as a lament, as though we *might* treat models as maps if only (at least) we attended to their domains of usefulness. Yet if Cartwright is correct, abstract concepts such as "rationality" or "institution" have no pre-established domain. Instead, we establish and re-establish their domains by looking to see what happens as we seek to extend or refine them in various ways. And, on Cartwright's approach, that is what we use interpretive models to do.

Leamer (2012, 16-7) suggests that "it seems better to say that we are writing mathematical novelettes, not making maps." This is what Cartwright does when she argues that we use models as we use fables. For present purposes, several aspects of the comparison between models and fables seem relevant. The models-as-fables view is less concerned with what a model *is* and more concerned with what we *do* with models. In large part this is because, like models, fables are exceedingly various: "what makes a fable is not the form of the material, but its orientation and the use made of it" (Blackham 1985, 136, 202-8). At bottom "a fable is basically a metaphor" albeit one that operates as "a narrative device, to provoke and aid concrete thinking, focused on some general matter of concern" (Blackham 1985, xi, xvii).[13] This, a specification of how abstract and concrete are related, is what preoccupies Cartwright (1999, 338-39). It is what is at stake in the way stories figure in model-driven inquiry.[14] Understanding how the narrative structure of the fable works to "fit out" a more abstract moral or lesson helps clarify how models mediate between our theories and the world.

The formal constraints on what counts as a fable are indeed loose. Consider just one frequent preconception. The characters who populate many fables are animals, chosen because they capture some particular, commonly recognized behavioral characteristics. One might think here of the various characters in well-known fables from Aesop through *Animal Farm.* But this conceit is not essential to the form. As Cartwright (1999, 42-3) notes the agents in a fable might just as well be "stereotypical persons." The sociopaths who populate conventional game theoretic models, those incessantly calculating, relentlessly self-interested players who care for the welfare of others or for the truth solely insofar as those niceties bear on their own payoffs, exemplify such creatures.

Suppose we grant that the models-as-fables view captures what we do with models. The obvious question is 'How do models work?" The answer is quite straightforward. As fables, models work in the translation of abstract and concrete, general and particular. "Thus, a fable is a story that represents a particular case that is immediately recognizable as a general principle" (Blackham 1985, 102). The broader or more enduring principle consists of the truths the fable reveals: the lesson, message or moral it conveys. The fable reveals truths "not sufficiently recognized" and does so not by argument but by showing or picturing or demonstrating what the general means in



https://doi.org/10.1017/S1537592720003473 Published online by Cambridge University Press

some particular instance (Blackham 1985, 175). But because it retains this connection to the general it is not merely an illustration.[15] A fable is used to provoke thought. Rather than instruct us what to think in some didactic way it suggests—or reminds us of—what we might and need to think about (Blackham 1985, xi, 223).

Myerson's (2007) discussion of game theoretic models of deterrence makes clear this commonality between models and fables. He depicts models as stories people use to make sense of problems that we encounter in the world. Specifically, he insists deterrence models offer broad "lessons" about the need to balance force and restraint in conflict situations. The difficulty is that, like models more generally, models of strategic deterrence are not vending machines. They do not dispense factory-wrapped confections. They prompt reflection. They call for judgement about what deterrence consists of, in this or that particular context. Myerson's exposition, in other words, is an exercise in practical reason, not coincidently addressed to active-duty military officers in hopes of imparting lessons that will deepen their appreciation of the strategic circumstances they will confront.

The deterrence models as Myerson interprets them hardly are idiosyncratic. His exposition follows in the footsteps of Schelling (1960) whose aim was to expand the range of possibilities available to Cold War military strategists (Myerson 2004; 2009). His lesson? That conflict is not zero sum but allows for a range of accommodations and many varieties of coordination that help parties avoid self-destructive conflagration. But even Schelling is hardly a unique example here. Humphreys (2017) offers a compendium of four dozen canonical "political games" from each of which he derives a lesson (he calls them "principles") of just the same sort. Models work as fables in a generalized fashion (Johnson 2014, 2019).

This seemingly raises a difficulty. In retrospect Cartwright (2010, 21-2) worries that some models do not convey a single readily identifiable lesson or moral and, insofar as she believes a fable must do so, she fears the models-as-fables view breaks down. She is especially concerned that the different lessons a model might convey "can point in different directions," and hence underwrite conflicting conclusions. Her concern, as it turns out, is needless because it is characteristic of fables that they need not be unambiguous in the way she believes.

Blackham anticipates and downplays the worry Cartwright articulates. As already noted, it is central to the way fables work that whatever moral they communicate prompts thought. Rather than telling us what to think, fables remind us what it is important to consider or to think about as we encounter problems in the world. What they offer is: "A message; no official message, no esoteric message, an uncertain message, an enigmatic message, what the reader tells him/herself; how unsatisfactory. Not really, because we are shown something every time" (Blackham 1985, 248). Thus, in order to successfully show what it does, to reveal truths, a fable must enlist its audience as collaborators. The same is true of models. They demand that those to whom they are addressed actively partake in an interpretive enterprise (Rubinstein 1991).[16] This, in part, is where the contested, argumentative nature of any research tradition is grounded.

## Rochester PPT: Fables of Rationality and Institutions

Cartwright illustrates her brief for the models-as-fables view with the concept of force as it is used in classical mechanics. In light of my extended discussion of Rochester-based PPT I instead focus on the abstract concepts of rationality and institution. I have argued that the most decisive contributions to that research tradition depart from the standard rationale. It remains to be seen how those same contributions conform to the models-as-fables view.

In the abstract, the concept of rationality is nearly empty. "All it means is that human behavior is mostly goal directed, often in a fairly consistent manner, in many important classes of social situations" (Harsanyi 1986, 84). To be useful for normative or explanatory—let alone predictive—purposes it is crucial to render this abstract concept more concrete. That is what we use game theoretic models to do.

Advocates of the standard rationale may find this claim preposterous. But consider two early characterizations of why we use game theory in social science.

> In analyzing games, the theory *does not* assume rational behavior; rather it attempts to determine what 'rational' can mean when an individual is confronted with the problem of optimal behavior in games and equivalent situations. (Morgenstern, 1968, 62; emphasis original)

> Game theory is the formal study of the rational, consistent expectations that participants can have about each other's choices… . It is, though, abstract and deductive, not the empirical study of how people make decisions but a deductive theory about the conditions that their decisions would have to meet in order to be considered "rational," "consistent," or "noncontradictory." Of course, defining "rational," "consistent," or "noncontradictory" for interdependent decisions is itself part of the business of game theory. (Schelling 1984, 215)

Here Morgenstern and Schelling second Cartwright's characterization of interpretive models. Each claims that game theoretic models are useful for conceptual purposes, for specifying what rationality consists in in some relatively particular, concrete circumstance or other. This in turn means, as Morgenstern insists, that what is distinctive about game theory is *not* that it "assumes" rationality and deduces predictions from that assumption. Instead it takes as background something like the abstract concept of rationality that Harsanyi articulates and allows us to interpret what that means within the concrete parameters



https://doi.org/10.1017/S1537592720003473 Published online by Cambridge University Press

of particular models. On this view particular game theoretic models supply us with working conceptions of rationality. In so doing they afford the technology for rationalizing behavior in various relatively concrete strategic settings and thereby for exploring the domain of our abstract concept of rationality.[17]

On this account game theory defines what counts as rational action in the quite specific context that any model captures. The exemplars of the Rochester PPT work on dimensionality and institutions I canvassed earlier use models in just the same way. That work constitutes an extended argument over what counts as rationality in politics. That argument revolves not, as the standard rationale suggests, around generating testable empirical predictions but around the task of provoking us to think. This is perhaps clearest in Riker (1980) who draws sweeping morals from the so-called "chaos" models and the institutionalist response to them.[18] It is important to recall, however, that McKelvey (1976; 1979) invites the sorts of inference Riker advances. In fact, he explicitly draws several morals or lessons from his technical proofs. Most obviously, McKelvey demonstrates that under the impossibly sparse conditions he specifies, majority rule is seriously ill-defined. Conversely, he concludes that, as Shepsle (1979, 53) puts it in his initial effort to characterize SIE and specify the conditions that might sustain them, "institutional details matter." Finally, McKelvey underscores how his technical results prompt us to think about the political importance of agenda control and the dangers of manipulation to which it gives rise. These conceptual "truths" are what emerge from the initial episode in the Rochester PPT models of disequilibrium and institutions (Ferejohn 2007).

We can say much the same about the way Calvert punctuates this intellectual argument a decade and a half later. His animating premise is a lament: "It is a measure of the importance of the idea of institutions in social science that the term often goes without an explicit definition in many studies to which it is central" (Calvert 1995a, 216). He then advances a definition of "institution" to compete with conceptions that "treat institutions as constraints on behavior or 'rules of the game'" (Calvert 1995b, 58). As he puts it: "The institution is just an equilibrium." And again: "Institution is just a name we give to certain parts of certain kinds of equilibria" (Calvert 1995b, 73,74). He is, in other words, defining "institution" by establishing what the concept means in linked series of particular settings. This is primarily a conceptual undertaking.[19] That said, Calvert (1995b 75) is correct to claim that his definition affords "a useful tool to guide empirical study of institutions" just insofar as we can insert it into the substantive claims we make in our theories about the world. In other words, far from generating predictions in the way defenders of the standard rationale imagine, Calvert offers conceptual resources for talking about both the persistence of institutional arrangements and the consequences they generate.

Calvert's approach to institutions rationalizes the behavior of the losers whom Riker expects invariably will conspire to alter institutional arrangements that work to disadvantage them. Calvert shows why, despite their discontent, losers might well *not* work to transform institutional arrangements. He does so by underscoring the obstacles to coordinating a successful effort to alter or replace extant arrangements. The reasoning is simple. If, in addition to inducing equilibrium outcomes as Shepsle shows, institutions are themselves equilibrium outcomes in an underlying repeated game, those disadvantaged in institutional equilibria cannot, under specifiable conditions and given their expectations about how others will act, improve their lot by departing from equilibrium behavior. Riker's objection notwithstanding, Calvert provides a way of thinking about how institutions persist and structure ongoing interactions among rational agents. Like earlier installments in the Rochester PPT argument over disequilibrium and institutions in politics, his contribution is conceptual. Like those earlier installments, Calvert's models are best understood as fables.

Blackham sketches the relevant implication succinctly.

> Fables are, like science, concerned with connections, with what goes with what and what follows what, but at a general level not amenable to reduced and refined statements testable by quantifiable evidence; nor are their "truths" merely loose empirical generalizations. They are concerned with connections and conditions… . There are truths of this kind which, when brought to notice are seen as evident, with recognition of the importance of their bearings. They are both undeniable and arguable; to deny them is to quibble, but to admit them is the first word only. (Blackham 1985, 252-3)

In social science, formal models are like fables. We use them as tools to define concepts central to our theories about the world. We use them to inform our ongoing conversation about common problems.

## Conclusion: Using Models-as-Fables

The case I advance reconstructs our understanding of how political scientists use formal models. It urges that we abandon the standard rationale and embrace instead a models-as-fables interpretation. How might we understand the force of this argument? Laudan (1996, 142-53) suggests criteria by which the sort of tale I tell should be assessed. It must, first, show that extant understandings of inquiry—in this case, what I call the standard rationale—are faulty. I have done that by establishing that the standard rationale is, on its own terms, of scant use in grasping the seminal work of PPT that emerged from Rochester from the late 1970s through well into the current century. None of the key developments in that research tradition generate predictions that might be tested. The argument between McKelvey, Schofield,



https://doi.org/10.1017/S1537592720003473 Published online by Cambridge University Press

Shepsle, Riker, and Calvert therefore *cannot* conform to the standard rationale. Subsequent efforts to reconcile PPT with the standard rationale by showing how formal models generate predictions have proven unsuccessful.

Second, any tale like mine also must, Laudan suggests, preserve the major accomplishments and crucial contributors recognized in extant accounts of the research tradition. Setting aside the prediction and testing understanding embodied in the standard rationale in no way calls into question the value of the models central to the Rochester analysis of disequilibrium and institutions in politics. I have simply offered more perspicuous reasons for appreciating those models and the arguments they sustain. In these two ways, my reconstruction acknowledges what Laudan terms "the rational weight of the past" in ongoing arguments over the contours and content of any research tradition. What fails is a particular interpretation (the standard rationale) of Rochester PPT, not the lessons or truths its most important models offer.

At this point, even a sympathetic reader might be left with a question—"What is it you hope to accomplish?" Indeed, I anticipate the plaintive voices—"Why do we need to worry about the matters you raise? Can't we please just get on with it?" where, by "it" they mean *real* research. I sympathize with the impatience. But my response would be that any claim regarding what counts as real research—especially for those who place a premium on the "rigor," "precision" and "clarity" of their arguments—must be sustained by some persuasive story of the sort I offer. The impatient reader owes us some such tale. I offer strong reasons to set aside the standard rationale and sketch a persuasive alternative account of why political scientists make and use formal models. The burden falls to impatient readers to either redeem the standard rationale or advance another, more compelling, alternative to the account I offer. I hope mostly to provoke such readers to accept that burden.

## Notes

1 See Myerson (2004, 102; 2007, 4-6,23-4; 2013, 305-6) and Rubinstein (2006, 881-2; 2012, 16-18,148,182; 2013).

2 Clarke and Primo 2012 and Leamer 2012 offer useful discussions of empirical models and how they relate to formal or theoretical models.

3 Empirical performance is crucial to our assessment of social scientific *theories.* No one denies that. Nothing I say implies otherwise. My claim is that empirical performance is a mistaken criterion for assessing formal *models* of the sort political scientists commonly use.

4 See e.g., Aldrich and Alt 2003, 310; Aldrich et al. 2008, 835; Granato and Scioli 2004, 314; Granato et al. 2010, 785; Monroe et al. 2015, 72-3; Bräuninger and Swalve 2020.

5 Critics of PPT also adopt this tendentious view; Mackie 2003, 174.

6 Austen-Smith and Banks (1998, 270–71; 1999, 84) and Cox and Shepsle (2007) both endorse this interpretation.

7 Knight 1995 elaborates what is at stake here by specifying the mechanisms that animate competing rational choice theories of institutional emergence and change.

8 One could hardly ask for a clearer statement of the standard rationale than Diermeier and Krehbiel (2003,127-9) offer. Diermeier reiterates this exact same characterization a decade later (2014a).

9 They depict Calvert (1995a, b) as instantiating their institutionalism as methodology view even as he departs from the standard rationale in the way I suggest.

10 Compare Austen-Smith and Banks 1998, 261, and Kreps 1990, 1.

11 To say of a concept like rationality or institution "that it is abstract is to point out that it *always* piggy-backs on more concrete descriptions" specified in some interpretive model or other; Cartwright 1999, 45, emphasis added.

12 Political theorists distinguish between general and abstract *concepts* and more particular and concrete *conceptions*; Rawls 1971, 5,9-10 and Dworkin 1986, 70-2.

13 It is a commonplace that "Models, and the metaphorical processes underlying them, are an integral part of science" (Bicchieri 1988, 108). That said, this leaves open the character of the "metaphorical processes" involved.

14 "In economists' use of models, there is always an element of interpretation: the model *always* tells a story" (Gibbard and Varian 1978, 666, emphasis added). Compare, for example, Leamer (2012, 18-24) on the distinction between the "mathematics" of a model and its "message." Compare too, Myerson (2007, 4-6) on the relationship of mathematics and narrative in game theory.

Thus, "at the frontier of economic argument, where we are arguing about what is important and what is not, the economist will need … a story. At the end of the mathematics she will ask insistently *why.* In seminars on economics the question 'What is your story?' is an appeal for a lower level of abstraction, closer to the episodes of human life"; McCloskey 1998, 15. Fables are stories stripped of many literary niceties (e.g. character development) and driven primarily by plot; Blackham 1985, 195.

15 "The message is not delivered … it is embodied. It is in this sense that the fable is a conceptual artefact which remains to be used. Interplay continues between the thought provoked and the representation that provokes and aids it"; Blackham 1985, xviii-xix.



https://doi.org/10.1017/S1537592720003473 Published online by Cambridge University Press

16 Thus, Humphreys does not simply offer a digest of canonical models. He encourages readers to explore each model and the refined lesson or principle he derives from it in a refreshing way. "Use them," he implores, "as a starting point for conversation, not as problem sets"; (Humphreys 2017, xxvi.

17 "What is the relation between a reason and an action when the reason explains the action by giving the agent's reason for doing what he did? We may call such explanations rationalizations, and say the reason *rationalizes* the action"; Davidson 1980, 4.

18 For Riker, when we explore the "devastating" implications of the McKelvey-Schofield results we ask a properly "philosophical question"; 1980, 441, 442. He draws a moral from those results: "And what we have learned is simply this: Disequilibrium, or the potential that the status quo be upset, is the characteristic feature of politics"; Riker 1980, 443.

19 It is *conceptual* insofar as it involves identifying and examining plausible causal mechanisms that commonly are unobservable. After all: "We cannot see, feel, touch, or even measure institutions"; North 1990, 107.

## References


Aldrich, John. 2011 [1995]. *Why Parties? A Second Look*. Chicago: University of Chicago Press.

Aldrich, John, and James Alt. 2003. "Introduction to the Special Issue." *Political Analysis* 1(4): 309–15.

Aldrich, John, James Alt, and Arthur Lupia. 2008 "The EITM approach: Origins and Interpretations." In *The Oxford Handbook of Political Methodology*, ed. Janet M. Box-Steffensmeier, Henry E. Brady, and David Collier. Oxford, UK: Oxford University Press.

Aldrich, John, and Arthur Lupia. 2011. "Experiments and Game Theory's Value to Political Science." In *Cambridge Handbook of Experimental Political Science*, ed. James Druckman, Donald P. Green, James H. Kuklinski, and Arthur Lupia, 89–101. Cambridge: Cambridge University Press.

Amadae, S.M., and Bruce Bueno de Mesquita. 1999. "The Rochester School: The Origins of Positive Political Theory." *Annual Review of Political Science* 2: 269–95.

Austen-Smith, David, and Jeffrey Banks. 1998. "Social Choice Theory, Game Theory, and Positive Political Theory." *Annual Review of Political Science* 1:259–87.

——. 1999. *Positive Political Theory I*. Ann Arbor: University of Michigan Press.

——. 2005. *Positive Political Theory II*. Ann Arbor: University of Michigan Press.

Bicchieri, Cristina. 1988. "Should a Scientist Abstain from Metaphor?" In *The Consequences of Economic Rhetoric*, ed. Arjo Klamer, Donald McCloskey, and Robert Solow. New York: Cambridge University Press.

Blackham, H.L. 1985. *The Fable as Literature*. London: Athelone.

Bräuninger, Thomas, and Tilko Swalve. 2020. "Theory Building for Causal Inference: EITM Research Projects." In *SAGE Handbook of Research Methods in Political Science and International Relations*, ed. Luigi Curini and Robert Franzese. Thousand Oaks, CA: Sage.

Bueno De Mesquita, Bruce, and Kenneth Shepsle. 2001. "William Harrison Riker, 1920–1993: A Biographical Memoir." *Biographical Memoirs*, Vol. 79. Washington, DC: National Academy Press.

Calvert, Randall. 1995a. "The Rational Choice Theory of Social Institutions." In *Modern Political Economy*, ed. J. Banks and E. Hanushek. Cambridge: Cambridge University Press.

——. 1995b. "Rational Actors, Equilibrium and Social Institutions." In *Explaining Social Institutions*, ed. J. Knight and I. Sened. Ann Arbor: University of Michigan Press.

Cartwright, Nancy. 1999. *The Dappled World: A Study of the Boundaries of Science*. New York: Cambridge University Press.

——. 2010. "Models: Parables v. Fables." In *Beyond Mimesis and Convention.* Boston Studies in the Philosophy of Science, Vol. 262. Ed. Roman Frigg and Mathew Hunter. New York: Springer.

Clarke, Kevin, and David Primo. 2012. *A Model Discipline*. Oxford: Oxford University Press.

Contessa, Gabrielle. 2011. "Scientific Models and Representation." In *Continuum Companion to the Philosophy of Science*, ed. Steven French and Juha Saatsi. New York: Continuum International Publishing,

Cox, Gary. 1999. "The Empirical Content of Rational Choice Theory." *Journal of Theoretical Politics* 11(2): 147–69.

Cox, Gary, and Kenneth Shepsle. 2007. "Majority Cycling and Agenda Manipulation." In *Positive Changes in Political Science*, ed. John Aldrich, James Alt, and Arthur Lupia. Ann Arbor: University of Michigan Press.

Davidson, Donald. 1980. *Essays on Actions and Events*. Oxford: Oxford University Press.

De Marchi, Scott. 2005. *Computational and Mathematical Modeling in the Social Sciences*. New York: Cambridge University Press.

Diermeier, Daniel. 2014a. "Positive Political Theory." In *The Blackwell Encyclopedia of Political Thought*, ed. Michael Gibbons. Hoboken, NJ: Wiley-Blackwell.

——. 2014b. "Formal Models of Legislatures." In *The Oxford Handbook of Legislative Studies*, ed. Shane Martin, Thomas Saalfeld, and Kaare Strøm. Oxford: Oxford University Press.




https://doi.org/10.1017/S1537592720003473 Published online by Cambridge University Press

Diermeier, Daniel, and Keith Krehbiel. 2003. "Institutionalism as Methodology." *Journal of Theoretical Politics* 15(2): 123–44.
Dowding, Keith. 2016. *The Philosophy and Methods of Political Science*. London: Palgrave Macmillan.
Dowding, Keith, and Charles Miller. 2019. "On Prediction in Political Science." *European Journal of Political Research* 58(3): 1001–18.
Dworkin, Ronald. 1986. *Law's Empire*. Cambridge, MA: Harvard University Press.
Elster, Jon. 2007. *Explaining Social Behavior*. New York: Cambridge University Press.
Ferejohn, John. 2007. "McKelvey's Democracy." In *Positive Changes in Political Science*, ed. John Aldrich, James Alt, and Arthur Lupia. Ann Arbor: University of Michigan Press.
Frigg, Roman, and Stephan Hartmann. 2012. "Models in Science." *Stanford Encyclopedia of Philosophy.* (http://plato.stanford.edu/archives/fall2012/entries/models-science/).
Gibbard, Allan, and Hal Varian. 1978. "Economic Models." *Journal of Philosophy* 75 (11): 664–77.
Giere, Ronald. 2004. "How Models Are Used to Represent Reality." *Philosophy of Science* 71:742–52.
Giere, Ronald, John Bickle, and Robert Mauldin. 2005. *Understanding Scientific Reasoning*. Belmont, CA: Wadsworth.
Granato, Jim, Melody Lo, and M.C. Sunny Wong. 2010. "A Framework for Unifying Formal and Empirical Analysis." *American Journal of Political Science* 54(3): 783–97.
Granato, Jim, and Frank Scioli. 2004. "Puzzles, Proverbs, and Omega Matrices: The Scientific and Social Significance of Empirical Implications of Theoretical Models (EITM)." *Perspectives on Politics* 2(2): 313–23.
Hacking, Ian. 1983. *Representing and Intervening*. New York: Cambridge University Press.
Harsanyi, John. 1986 [1977]. "Advances in Understanding Rational Behavior." In *Rational Choice*, ed. Jon Elster. New York: New York University Press.
Hausman, Daniel. 1992a. *Essays on Philosophy and Economic Methodology*. Cambridge University Press.
——. 1992b. *The Inexact and Separate Science of Economics*. Cambridge University Press.
Humphreys, Macartan. 2017. *Political Games: Mathematical Insights on Fighting, Voting, Lying, and Other Affairs of State*. New York: W.W. Norton.
Johnson, James. 2014. "Models Among the Political Theorists." *American Journal of Political Science* 58: 547–60.
Johnson, James. 2019. "Formal Models in Political Science: Conceptual, Not Empirical." *Journal of Politics* 81(1): e6–e10.
King, Gary, Robert O. Keohane, and Sidney Verba. 1994. *Designing Social inquiry: Scientific Inference in Qualitative Research*. Princeton, NJ: Princeton University Press.
Kitcher, Phillip. 1993. *The Advancement of Science*. Oxford University Press.
Knight, Jack. 1995. "Models, Interpretations and Theories." In *Explaining Social Institutions*. Ed. J. Knight and I. Sened. Ann Arbor: University of Michigan Press.
Krehbiel, Keith. 1988. "Spatial Models of Legislative Choice." *Legislative Studies Quarterly* 13(3): 259–319.
Krehbiel, Keith, Adam Meirowitz, and Jonathan Woon. 2005. "Testing Theories of Lawmaking." In *Social Choice and Strategic Decisions*, ed. David Austen-Smith and John Duggan. New York: Springer.
Kreps, David. 1990. *Game Theory & Economic Modelling*. Oxford, UK: Oxford University Press.
Laudan, Larry. 1977. *Progress and Its Problems*. Berkeley: University of California Press.
——. 1996. *Beyond Positivism and Relativism*. Boulder, CO: Westview Press.
Leamer, Edward. 2012. *The Craft of Economics*. Cambridge, MA: MIT Press.
MacIntyre, Alasdair. 1977. "Epistemological Crises, Dramatic Narrative and The Philosophy of Science." *The Monist* 60(4): 453–72.
Mackie, Gerry. 2003. *Democracy Defended*. New York: Cambridge University Press.
McCarty, Nolan, and Adam Meirowitz. 2007. *Political Game Theory: An Introduction*. New York: Cambridge University Press.
McCloskey, Deirdre. 1998 [1985]. *The Rhetoric of Economics*. 2nd ed. Madison: University of Wisconsin Press.
McDermott, Rose. (2002). "Experimental Methods in Political Science." *Annual Review of Political Science* 5(1): 31–61.
McKelvey, Richard. 1976. "Intransitivities in Multidimensional Voting Models and Some Implications for Agenda Control." *Journal of Economic Theory* 12(3): 472–82.
——.1979. "General Conditions in Formal Voting Models." *Econometrica* 47(5): 1085–12.
Monroe, Burt, Jennifer Pan, Margaret E. Roberts, Maya Sen, and Betsy Sinclair. 2015. "No! Formal Theory, Causal Inference and Big Data are Not Contradictory." *PS: Political Science & Politics* 48(1): 71–74.
Morgan, Mary, and Margaret Morrison, eds. 1999. *Models as Mediators*. New York: Cambridge University Press.
Morgenstern, Oskar. 1968. "Game Theory I: Theoretical Aspects." In *International Encyclopedia of the Social Sciences*. Vol. 6. Ed. David Sills. New York: Macmillan.



https://doi.org/10.1017/S1537592720003473 Published online by Cambridge University Press

Morton, Rebecca. 1999. *Methods & Models. A Guide to the Empirical Analysis of Formal Models in Political Science*. New York: Cambridge University Press.
——. 2009. "Formal Modeling and Empirical Analysis in Political Science." In *Methoden der Vergleichenden Politik- und Sozialwissenschaft*, ed. Susanne Pickel, Gert Pickel, Hans-Joachim Lauth, and Detlef Jahn. New York: Springer Verlag.
Morton, Rebecca, and Kenneth Williams. 2008. "Experimentation in Political Science." *The Oxford Handbook of Political Methodology*, ed. Janet M. Box-Steffensmeier, Henry E. Brady, and David Collier. Oxford, UK: Oxford University Press.
Myerson, Roger. 2004. "Justice, Institutions and Multiple Equilibria." *Chicago Journal of International Law* 5:91–107.
——. 2007. *Force and Restraint in Strategic Deterrence: A Game-Theorist's Perspective*. Strategic Studies Institute. Carlisle, PA: U.S. Army War College.
——. 2009. "Learning from Schelling's *Strategy of Conflict*." *Journal of Economic Literature* 4794): 1109–25
——. 2013. "Fundamentals of Social Choice Theory." *Quarterly Journal of Political Science* 8(3): 305–37.
North, Douglass. 1990. *Institutions, Institutional Change and Economic Performance*. New York: Cambridge University Press
Norton, Anne. 2004. *95 Theses on Politics, Culture & Methods*. New Haven, CT: Yale University Press.
Ordeshook, Peter. 1986. *Game Theory and Political Theory*. New York: Cambridge University Press.
——. 1992. *A Political Theory Primer*. London: Routledge.
Ostrom, Elinor. 1990. *Governing the Commons*. New York: Cambridge University Press.
——. 1998. "A Behavioral Approach to the Rational Choice Theory of Collective Action." *American Political Science Review* 92(1): 1–22.
Patty, John, and Elizabeth Maggie Penn. 2014. *Social Choice and Legitimacy*. New York: Cambridge University Press.
Rawls, John. 1971. *A Theory of Justice*. Cambridge, MA: Harvard University Press.
Riker, William. 1977. "The Future of a Science of Politics." *American Behavioral Scientist* 21(1): 11–38.
——. 1980. "Implications of the Disequilibrium of Majority Rule for the Study of Institutions." *American Political Science Review* 74(2): 432–46.
——. 1990. "Political Science and Rational Choice." In *Perspectives on Positive Political Economy*, ed. James Alt and Kenneth Shepsle. New York: Cambridge University Press.
Riker, William, and Peter Ordeshook. 1973. *An Introduction to Positive Political Theory*. Upper Saddle River, NJ: Prentice Hall.
Rosenberg, Alexander. 2005. *Philosophy of Science*. 2nd ed. London: Routledge.
Rubinstein, Ariel. 1991. "Comments on the Interpretation of Game Theory." *Econometrica* 59(4): 909–24.
——. 2006. "Dilemmas of an Economic Theorist." *Econometrica* 74(4): 865–883.
——. 2012. *Economic Fables*. Cambridge, UK: Open Book Publishers.
——. 2013. "How Game Theory Will Solve the Problems of the Euro Bloc and Stop Iranian Nukes." *Frankfurter Allgemeine,* March 27. (https://www.faz.net/aktuell/feuilleton/debatten/game-theory-how-game-theory-will-solve-the-problems-of-the-euro-bloc-and-stop-iranian-nukes-12130407.html).
Satz, Debra, and John Ferejohn. 1994. "Rational Choice and Social Theory." *Journal of Philosophy* 91(2): 71–87.
Schelling, Thomas. 1960. *The Strategy of Conflict*. Cambridge, MA: Harvard University Press.
——. 1978. *Micromotives & Macrobehavior*. New York: W.W. Norton.
——. 1984 [1967]. "What Is Game Theory?" In *Choice and Consequence*. Cambridge, MA: Harvard University Press.
Schofield, Norman. 1978. "Instability of Simple Dynamic Games." *Review of Economic Studies* 45(3): 575–94.
——. 2001. "Constitutions, Voting and Democracy: A Review." *Social Choice & Welfare* 18:571–600.
Shepsle, Kenneth. 1979. "Institutional Arrangements and Equilibrium in Multidimensional Voting Models." *American Journal of Political Science* 23(3): 27–59.
——. 1986. "Institutional Equilibrium and Equilibrium Institutions." *Political Science: The Science of Politics*. Edited by Herbert Weisberg. Bronx, NY: Agathon Press.
——. 2006. "Old Questions and New Answers About Institutions: The Riker Objection Revisited." In *The Oxford Handbook of Political Economy*, ed. Barry Weingast and Donald Wittman. Oxford: Oxford University Press.
——. 2010. *Analyzing Politics: Rationality, Behavior and Institutions*. 2nd ed. New York: W.W. Norton.
Woon, Jonathan. 2012. "Laboratory Tests of Formal Theory and Behavioral Inference." In *Experimental Political Science: Principles & Practices*, ed. Bernhard Kittel, Wolfgang Luban, and Rebecca Morton. London: Palgrave.



https://doi.org/10.1017/S1537592720003473 Published online by Cambridge University Press