

Game Theory, Information, and Deliberative Democracy

Dimitri Landa New York University
Adam Meirowitz Princeton University

We contend that, with a suitably broad notion of rationality and a diverse set of motivations, the game-theoretic tradition is particularly well suited for generating insights about effects of deliberative institutions and that progress in the development of deliberative democratic theory hinges on making proper sense of the relationship between game-theoretic and normative theoretic approaches to deliberation. To advance this view, we explore the central methodological issues at the core of that relationship and address the arguments raised against the relevance of game-theoretic work on deliberation. We develop a framework for thinking about the differences in how the normative and the game-theoretic approaches frame and answer questions about deliberation and articulate an approach to a deliberative democratic theory that builds on the strengths of both of these theoretic traditions, properly informed by empirical scholarship.

Democracy is minimally defined as a form of governance in which policy decisions are made by a majority vote of the citizens. While useful as a rough way of classifying polities, this definition is, on closer examination, both ambiguous and radically incomplete. The main thrust of the critique of minimal democracy developed in contemporary democratic theory is that voting is not the best, or at least not the only, political mechanism for ensuring that policy decisions conform to the interests of the citizens. A key political mechanism that also serves that role and that is missing from the minimalist view of democracy is deliberation, and the appreciation of the effects of this mechanism is changing the way scholars of democracy think about democratic institutions.

In revealing correct, fuller, or simply better organized information, deliberation provides an opportunity for participants to arrive at more considered judgments themselves and to affect collective decision making by influencing the judgments of others. Its consequences may affect what happens in a voting booth or in a legislative or a judicial chamber, or in the way

we approach a personal moral conundrum. A political decision-making process that fails to create the opportunity for or to take advantage of these benefits of deliberation is bound to raise questions about the legitimacy of the resulting outcomes (Cohen 1996; Habermas 1996; Manin 1987).

Apart from the direct effects of better information, deliberation contributes to the legitimacy of policy choices and of the underlying political institutions in a number of other ways. It can raise the sense of political autonomy and of the effective fairness of policy choices, enable a better assessment of fellow citizens' motives with respect to a given political choice, and even encourage the production of other-regarding motives (Elster 1995). It may also increase the stability of collective choice by reducing the number of issue dimensions and introducing more structure into individual preferences (Dryzek and List 2003; Knight and Johnson 1994). But to have these effects, deliberation must bring about some kind of learning that changes participants' preferences over choices.¹ At bottom, the transmission, processing, and aggregation of the information that forms the basis of individual and

Dimitri Landa is Assistant Professor of Politics, New York University, 19 West 4th Street, New York, NY 10012 (dimitri.landa@nyu.edu). Adam Meirowitz is Associate Professor of Politics, Princeton University, Corwin Hall, Princeton, NJ 08544 (ameirowi@princeton.edu).

The authors thank Scott Ashworth, Sandy Gordon, Cathy Hafer, David Stasavage, and participants at the 2006 Princeton Conference on Deliberative Democracy for stimulating discussions and comments on earlier drafts.

¹This is, arguably, true regardless of whether the implied transformation is of the participants' "primitive" or their "induced" preferences.

²See the instructive discussion of causal claims on behalf of deliberation in Fearon (1998).

³Throughout, we use the terms "normative theory" and "political philosophy" and their derivatives interchangeably.

American Journal of Political Science, Vol. 53, No. 2, April 2009, Pp. 427–444

collective decision making is the engine that drives the deliberative wheels.²

Although normative theorists have fashioned the research agenda of deliberative democratic theory,³ that same informational engine is the key motivation behind the now very extensive game-theoretic literature on communication. An emerging body of work in this literature focuses specifically on policymaking in deliberative institutions. One might anticipate, therefore, that there would be a great deal of interaction between the normative and the game-theoretic work on deliberation. Alas, that is not the case. While game-theoretic studies sometimes explicitly challenge specific assertions that emerge in the normative literature, game theorists tend to exert little effort toward making their results accessible to a less technical readership and rarely undertake the reconstruction of normative arguments in a manner consistent with the insights from the game-theoretic models. On the other hand, the normative literature on deliberation, with very few exceptions, takes essentially no account of the presence of the game-theoretic work on deliberation and ignores the fundamental incentive problems that surface in nearly all relevant game-theoretic studies.

The dissimilar styles of exposition and the high entry costs of each approach, surely, contribute to the failure to see eye-to-eye. But these reasons are, not surprisingly, rarely invoked self-consciously. The explicit reasons that appear *prima facie* more defensible and which appear in print are, in our reading, threefold: (1) it is unclear how to make sense of the analytical/structural relationship between these approaches: what kind of contribution, if any, can each approach make to the pursuit of the agenda set by the other? (2) the communication analyzed by game theorists is of a fundamentally different epistemic type, and the game-theoretic results are, therefore, largely irrelevant to deliberative democratic theory; and (3) the game-theoretic approach omits key social and philosophic determinants of deliberation, and its conclusions are therefore irrelevant because they are an artifact of these omissions. In one form or another these reasons have been used to justify the persistent divide between the two approaches.

Our goal in this article is to sketch what we take to be the most productive and defensible way to think about the relationship between these approaches and to make, against the critics, the case for the relevance of the existing game-theoretic analysis of deliberation to the development of deliberative democratic theory. We argue that the reasons for the mutual avoidance of the game-theoretic and the normative traditions are not supportable, and that the common view that game-theoretic scholarship of deliberation is aimed at undermining de-

liberative democracy is mistaken.⁴ While a considerable portion of our article is dedicated to clarifying the role of the game-theoretic approach, this position does not amount to the claim that it, unlike other approaches, gets it “right,” but rather that it is, alongside normative arguments, an essential element of the account of construction of deliberative democratic theory that we propose.

Before proceeding, we pause to note a third important body of scholarship on deliberation, which has emerged from the social psychology and experimental traditions. Contributions by Druckman (2004), Mutz (2006), Jackman and Sniderman (2006), and Taber and Lodge (2006) are recent examples of this work. Mendelberg (2002) provides an influential review of the classic work in this tradition, elucidating the ways that theory and data from social psychology challenge and reinforce aspects of the normative literature. Most recently, Mutz (2008) presents a sustained argument for the value of empirical evaluation of claims made by normative theories of deliberation. The connections between the game-theoretic and the psychological/experimental literature are regrettably weak (Lupia 2002). Although the focus of our arguments in what follows prevents us from providing a systematic treatment of those connections, we point, where appropriate, to intersections and potential complementarities between the respective research agendas.

Our mode of analysis is, in large part, reconstructive—we suggest ways of thinking about the approaches to deliberation that may differ from (though need not be at odds with) the explicit ways in which their respective partisans think about them, and we deliberately avoid a sustained survey of the results from the existing work in the interest of focusing on the broad-brush questions of underlying analytical relationships. Since our goal is to reach a broad audience, including those who have little or no familiarity with game-theoretic reasoning, we also adopt a nontechnical exposition and offer informal self-contained examples to illustrate our arguments.

In the next section we lay out the analytical details of our view of the relationship between normative and game-theoretic approaches to deliberation. The discussion in the remainder of the article addresses a number of key issues related to how and whether the game-theoretic analysis of deliberation furthers the research agenda implied by this view. To this end, the following section

⁴Both that such a view existed and that it was widespread among our colleagues came as a surprise to us. Our casual survey of scholars who work on game-theoretic models of deliberation revealed that it was quite the opposite of how they saw themselves. As our arguments below indicate, one of our primary goals is to articulate a defensible methodological apparatus for developing better deliberative democratic theory, not for convincing others that that theory should be rejected.

introduces some of the conceptual distinctions that have surfaced in the game-theoretic analysis and makes the case that the incentive problems that motivate it persist across the epistemic environments that have been the focus of deliberative democrats. The last substantive section builds on this case and responds to the critics who view the game-theoretic results as an artifact.

Decomposing Social Interactions

Deliberation-as-Behavior versus Deliberation-as-Environment

One might expect that a first step in advancing our understanding of deliberation would entail settling on the precise definition of this concept. Perhaps surprisingly, then, the general label of “deliberation” is used to refer to a broad range of phenomena that are sometimes only tangentially related. A somewhat rough but relatively faithful way of capturing the key underlying differences in formulating definitions of deliberative democracy turns on the difference between *behavior* and *environment*.⁵

For most, though not all, normative theorists, the theoretical enterprise focuses on specifying constraints on individual and group deliberative behavior (see, e.g., Bohman and Rehg 1997; Dryzek 2000; Gutmann and Thompson 1996). For this perspective, a stylized example is a claim along the lines of “deliberative democracy is a process by which individuals who are committed to offering justifications and seeking understanding actively and sincerely debate merits of policy alternatives in search of the policy that is most consistent with an idea of the common good.” Scholarship in this tradition is focused on the appropriate definition of a normative ideal of deliberative democracy. It advances this goal by presenting an account of ideal behavior by citizens, including what that behavior looks like in relation to particular social and political institutions and settings. Sometimes such accounts also concern outcomes (epistemic, distributive, or other) that may be expected to result from a deliberative democratic practice, but their consideration of outcomes is mainly a by-product of the focus on behavior. The underlying claim is that it is such and such ideal behavior that will bring about such and such outcomes. In short, this scholarship may be understood as treating deliberation and deliberative democracy as behavior—a profile of (normatively defensible) actions and choices on the part of

the citizens, or, equivalently, a profile of restrictions on what admissible behavior would be.

In contrast, the approach taken by game theorists involves defining deliberation not as a set of restrictions on behavior but rather as a set of restrictions on the environment in which participants interact. In this approach, a stylized example is the claim “deliberative democracy is an institution in which participants have the opportunity to make speeches prior to voting.” This claim, then, restricts attention to whatever decision making would be arrived at in the particular ways implied by the features of the environment (namely, by first exchanging speeches, and then casting ballots based on the judgments that, presumably, reflect the content of prior speech).

The game-theoretic approach involves a three-stage process. The first stage defines a game, which captures (a) the relevant choices that are understood to be available to the players (in models of deliberation, typically, what messages, if any, could be sent, and what decisions could be made after the exchange of messages), (b) what the players know about those choices, about each other, and about the deliberative interaction to which they are a party, and finally, (c) how attractive they would perceive the consequences of those choices to be if they knew everything that there was to know about them. The second stage specifies a solution concept, which embodies a set of assumptions about the behavioral agency (a general model of individual behavior that yields predictions about the actions or behavior of the participants when the environment of choice is specified) ascribed to the players in the model.⁶ Given the first two stages, the third stage is logically entailed: through well-defined techniques of

⁵We caution readers against associating the word *behavior*, as we use it, with its usage in describing the political behavior research tradition. Here we use the term “behavior” because it most appropriately captures what we believe is a key distinction: do we describe as deliberative *the way that individuals actually behave* or the institution or venue *within which they behave*?

⁶Work on deliberation as environment typically posits *an external set of behavioral assumptions* that have implications for how collective bodies behave in any environment. This enables the deductive determination of what deliberation will look like in a particular environment subject to the condition that the external assumptions are true. This approach suggests two directions for analysis—varying the external assumptions, holding fixed the description of the environment (e.g., Dickson, Hafer, and Landa 2008a; Hafer and Landa 2007), and varying the description of the environment, holding fixed the external assumptions (e.g., Austen-Smith and Feddersen 2005, 2006; Dickson, Hafer, and Landa 2008b; Meirowitz 2006, 2007). The analysis of deliberation as environment, thus, does not require that the external assumptions conform to a particular a priori fixed notion of rationality, though the debate about which behavioral agency is appropriate for which deliberative environments in the game-theoretic work is still nascent (Landa 2006). Ideally, progress would come by way of an iterative process of model formation, analysis, and testing leading to refinements. In this sense, our conceptualization of the deliberation-as-environment approach is entirely consistent with the work that draws on insights from the social psychology tradition (e.g., in the spirit of Miller and Davidson-Podgorny 1987, who provide microfoundations for understanding how the distribution of groups influences the possibility of cooperation). Armed with a refined behavioral agency, the analysis of how deliberative institutions influence behavior would then proceed as outlined above.

analysis, one can generate predictions about what types of behavior, with respect to the particular choices analyzed in the model, are and are not mutually consistent—that is, are or are not supportable by *equilibria* of the specified game. The key question that motivates the game-theoretic analysis is how policy selection is related to private information and preferences when participants engage in equilibrium behavior. Although this work aims to produce characterizations of behavior, it is important to see that what it has to say about deliberative behavior is induced by its focus on the nature of the deliberative environment. The analysis it delivers is the analysis of the properties of that environment, and it contributes to the study of deliberation insofar as that environment captures the essential institutional features of deliberative democracy. In short, this work treats deliberative democracy as an environment.

It bears emphasizing that the distinction between treating deliberation as behavior and treating it as an environment is a conceptual divide, rather than a sociological one separating normative from formal theorists. In this sense, our usage of the term “game-theoretic,” as opposed to “formal” is revealing. The critical contrast is not between formalized theory and nonformalized theory, but between the game-theoretic focus on the deliberative environment and the normative behavioral focus on the behavior without inducing it from the environment. Indeed, contributions of some scholars of deliberation who would identify themselves as normative theorists are, in the language of our conceptual formulation, best understood as treating deliberation as institution and so are in the spirit of the game-theoretic work (see, e.g., James 2003; Knight and Johnson 2007).

Democratic Ideals and Game-Theoretic Equilibria

As we noted above, the primary source of the deliberation-as-behavior approach has been the work of normative theorists. One may characterize a key contribution of this work as the articulation of axioms that are definitive of the democratic ideal and of its various conceptions—e.g., participatory, deliberative, epistemic, representative, direct, etc. Some of these axioms are procedural—i.e., descriptive of the process of deliberation and decision making, such as honest and open-minded participation, etc. (e.g., Cohen 1996; Gutmann and Thompson 1996). Other axioms are best understood as consequentialist—e.g., collective decision making selects a policy that is (more) correct, or carries with it minimal expected deviation from the policy that would have been chosen by a single omniscient decision maker with complete in-

formation (Estlund 1997). More generally, the axioms that form the basis of normative theorists’ descriptions of democratic ideals may be seen as the codifications of expectations with respect to both the aspects of deliberative environments and the behavioral choice in deliberation and voting.

Of course, normative theorists do not merely define the axioms; they aim also to make the case that, in relation to those axioms, some conceptions of democracy and some social institutions are better than others at capturing the democratic ideal. It is in connection with this step that we argue for the importance of a give-and-take with the game-theoretic analysis of deliberative environments. The comparison of the two approaches underscores the following key methodological point. Whereas the game-theoretic/deliberation-as-environment approach has an agreed-upon “machine” (or, more accurately, a small set of “machines”) for relating descriptions of the environment to descriptions of behavior and so for generating comparisons about how different descriptions of the environment might influence the nature of discourse and policymaking, the deliberation-as-behavior approach lacks such a device. Taken on its own, scholarship in this approach is not typically subject to explicit rules about “how people operate” (indeed, that is the point—since its goal is to produce normative claims for how people *should* operate). Each behavior-focused account of what deliberation means includes a potentially novel description of the rules of behavior and the rules for drawing inferences. The point is not that that approach necessarily does a worse job capturing deliberation than the game-theoretic one. It is, rather, that it has not yet equipped itself with an epistemic mechanism for discriminating between different accounts that instantiate it.

In seeking to satisfy axioms that capture the underlying aspects of the democratic ideal, we hope for what may be called *axiomatic consistency*. A key claim of the game-theoretic approach can be put as follows: an appropriately specified game-theoretic model allows us to gauge axiomatic consistency by ascertaining the satisfaction of *strategic behavioral consistency*. To explain what we mean by this, we first briefly sketch the key elements of the game-theoretic approach and then interpret them in relation to these notions of consistency.

In order for communication to do more than just allow participants to coordinate on a particular choice (that is, if deliberation is about convincing and/or being convinced by one’s interlocutors), one or more participants must be uncertain about some aspect of policy choice.⁷ The game-theoretic approach to modeling situations with

⁷See, though, Calvert (2006) for a model of cheap-talk coordination without policy uncertainty.

such uncertainty is to assume that some, and possibly all, participants are endowed with pieces of information that their counterparts do not know but would find relevant to the decision at hand.⁸ Other participants know that each participant might have some privileged information but do not know what that information is.

Consider the following example loosely modeled on a recent event—perhaps the last major policy debate in the United States. The United States is considering whether to engage in a military action to overthrow Saddam Hussein's regime in Iraq. The public and the members of the administration are divided into three groups. Two of these groups, the prowar camp and the antiwar camp, are composed of those whose minds are essentially made up. In contrast, members of the third, and pivotal, group are deeply ambivalent about the best way to proceed. Two key considerations may sway their opinion: the presence or absence in Iraq of weapons of mass destruction and the costs and the likelihood of establishing a stable, preferably democratic, regime to replace the dictatorial rule of Saddam Hussein. Some of the evidence pertaining to these considerations is public, some is classified, and members of the third group do not have direct access to some of the key classified information. They do not know what that information is, but they are aware of its existence, and can form beliefs about it on the basis of the information that is available to them.

Those preliminary beliefs enable them to make educated decisions in the absence of further information and to make sense of the signals or cues they may receive from others, but on their own, those beliefs are “noisy.” A key issue in game-theoretic models of policymaking is whether it is reasonable to expect those participants who possess valuable information to reveal it to others, and whether those others have good reasons to believe it. In the game of poker, which has some strategic properties in common with the example of the decision process regarding the invasion of Iraq, expecting truthful revelation of someone's face-down cards is unreasonable. We may observe small talk and gestures, but the savvy player will not take such communication at its face value—not the least, because she knows that competitive players go to great lengths to make sure that both verbal and nonverbal communication does not betray them. Whether a similar expectation is appropriate in the example of the consideration of military invasion in Iraq is an open question—it will depend on a long list of factors, including the values

of the members of the corresponding groups, the consequences of making inaccurate statements about what the classified information says, the precise manner in which the ultimate policy decision is to be made, etc.

The equilibrium notions that game theorists use to analyze such interactions are nothing more than descriptions of how the participants play the game. Such a description must satisfy the condition that, when the players share a common conjecture about how the game is being played, no player (conditional on any particular realization of her private information) has an incentive to deviate from the behavior that that description ascribes to her. In practice, an equilibrium concept may be somewhat more complicated because it also specifies the inferences that participants should make when particular speeches are observed. These inferences depend on the conjectured behavior (e.g., a conjecture that if the evidence for the existence of WMDs in Iraq is weak, then those who have access to it are unlikely to use phrases like “slam dunk” to describe the affirmative case), the underlying environment (e.g., the incentives for the competing intelligence agencies to reinforce or challenge each other's public pronouncements), and the assumption of behavioral agency being maintained. In other words, while participants may possess uncertainty about what others know, an equilibrium consists of a conjecture about the rules governing the behavior of individuals (e.g., always tell the truth, or tell all and only the truth under conditions x, y , or z and lie in some specified way otherwise), and in equilibrium, participants draw inferences from the speeches that are made, using that conjecture. Given the equilibrium conjecture, when participant A hears participant B say something like “going to war is a terrible idea,” she can form beliefs about what participant B knows about the relevant policy alternatives.

An equilibrium corresponding to a deliberative environment must satisfy three conditions: (1) Given the way that participants are forming beliefs based on communication, participants vote for the policies that they think they like best. (2) Given the ways that participants are behaving in the policy selection stage (which is dependent on how they are forming these beliefs based on communication), all of the participants have an incentive to communicate in the manner conjectured. (3) Given the way that participants are conjectured to behave at the communication stage, the beliefs that are formed are consistent with the postulated process.⁹ Taken together, these

⁸This uncertainty can range from the effects of policy choices and the identity of a true state of the world, to the logical (most efficient) way of organizing other information that is already available to the participants.

⁹Examples include belief updating consistent with Bayes Rule, in which case the equilibrium is sometimes termed the Perfect Bayesian Equilibrium, as well as inefficient updating, as in the models in Hafer and Landa 2005, 2007 and Ottaviani and Squintani (2006).

criteria of equilibrium behavior amount to the requirement of what we referred to above as *strategic behavioral consistency*—the mutual consistency of individual behavior in a game-theoretic equilibrium.

To make the implications of this requirement a bit clearer, we return to the poker example and ask what types of behavior can occur in equilibrium. Is it possible for there to be an equilibrium in which participants truthfully announce their face-down cards to each other? No; the benefits of successful bluffing ensure that such communication is not credible. Less obvious is the claim that it is not possible in equilibrium for all players with good hands to be quiet and all players with bad hands to be chatty. In such an equilibrium a player with a bad hand might have an incentive to be quiet to fool her opponents into believing that she has a good hand. Also players with good hands may chat in order to up the betting. As we discuss below, this type of reasoning has been applied to the study of deliberative policymaking.

We can now see how the game-theoretic analysis of strategic behavioral consistency sheds light on the question of axiomatic consistency. In effect, by analyzing the properties of the equilibria of the relevant game-theoretic model, we obtain propositions about the compatibility of various behavioral expectations, including axioms of truthful or complete information revelation by all participants, with each other and with the particular aspects of the deliberative environment that define the game.¹⁰ In the equilibrium analysis, this test may happen *ex post*: an axiom that is thought to ensure a particular consequence may, through the equilibrium analysis of deliberative environments that vary with respect to it, sometimes be discovered to imply altogether different, unanticipated consequences as well (or, put differently, be compatible or incompatible with a particular set of *ex post* expectations or axioms). In this way, recent game-theoretic work on deliberation has called into question the expectations associated with such axioms as the preference for participation and diversity of deliberative bodies (Meirowitz 2007), the equality of opportunity to make one's arguments heard (Hafer and Landa 2007), and the requirement of consensus in collective decision making (Austen-Smith and Feddersen 2006; Gerardi and Yariv 2006).

¹⁰Some game-theoretic work (e.g., Levy 2007; Meirowitz 2007; Stasavage 2007) treats a specific game and generates a precise description of "equilibrium play." Others (e.g., Austen-Smith and Feddersen 2006; Meirowitz 2006; Patty 2008) treat large classes of games, but generate somewhat less precise characterizations of deliberative behavior within a given entailed game.

The Three-Step Deliberative Democratic Theory

To the extent that democratic theory has practical aspirations, it seems difficult to argue that the mutual inconsistency of axioms that characterize normative conceptions of democracy is irrelevant. This suggests the value of developing scholarship on deliberative democracy that is responsive to the considerations of consistency offered by the game-theoretic analysis. In our view, scholarship of this form may be instructively conceived as consisting of the following three steps. The first step corresponds to the formulation of axioms, both procedural and consequentialist, in relation to the underlying political environment. The second step entails the analysis of axiomatic consistency within a corresponding game-theoretic model, including the consistency of the proposed axioms with the behavioral agency that is thought to characterize the agents operating in that deliberative environment. The third step in the construction of deliberative democratic theory closes the loop: it calls for a review of the normative conceptions and axioms with which the process began and sanctioning trade-offs where they are necessary.

The first step of our three-step account has been and, given its underlying basis in the normative theories of value, will likely remain the domain of normative theorists. But we see both the second and third steps as inherently requiring a sustained interaction between different approaches to analyzing deliberation and, in that, a departure from the present practice.

Consider the second step first. The analysis of axiomatic consistency, as we describe it above, is something to which game theory is particularly well suited. But, in a fundamental sense, game-theoretic analysis goes only where specifications of the underlying games take it. We argued above that a critical part of that specification comes from the normative analysis of axioms in the first step of our account. There are two further sources of that specification originating in empirical scholarship in social psychology, behavioral economics, and cognitive science. Much of that scholarship has focused on the analysis of behavioral agency. Responding to it, as some game theorists are now starting to do (see footnote 6 above), can lead to the revision of the assumptions underlying the equilibrium concepts in the game-theoretic analysis and, indeed, to generating different conclusions about what axioms are consistent. The game-theoretic analysis of axiomatic consistency may also be conditioned on assessments of how a given category of agents perceives features of the "deliberation games." It matters, for example, whether the agents are posited to be motivated by epistemic considerations or also, say, by a primitive

preference for conformity; whether the agents perceive the deliberative interaction as being relevant to later events that entail the imposition of negative sanctions contingent on their choices in deliberation, etc. The equilibria of the “deliberation games,” and so the judgments of axiomatic consistency they deliver, will differ depending on which preference is posited, what the game form that is understood to be commonly known to the players is, etc. Empirical analysis that sheds light on which models better explain the observed behavior in a given type of deliberative group ought to influence the choice of the model that is used to arrive at the judgments of axiomatic consistency (Dickson, Hafer, and Landa 2008b).¹¹

The third step presupposes a conversation between research traditions. On the normative-theoretic side, the conclusions of the second step raise the following question: if not all of the good things (expressed by the relevant axioms) can go together, and not all of those good things have the consequences that we thought them to have, what trade-offs between axioms are normatively justifiable?¹² For the game-theoretic and the empirical side, the normative-theoretic answers to this question pose a counterpart question: what are the conditions that maximize the feasibility of implementing the most defensible such trade-offs?

In addressing itself to this latter question, game-theoretic analysis becomes an essential tool of institutionally prescriptive normative theory. (The relationship between game theory and empirical analysis here is essentially that which we described in the second step of our account.) Assessing the mutual compatibility of behavioral axioms allows us to conduct comparisons of different deliberative environments by comparing properties of their equilibria. Because environments are, in part, defined by the particular institutions instantiating them, we

can thus arrive at a ranking of institutions in relation to various normative properties and, in so doing, ground the normative arguments regarding institutional choice. A central intuition behind this type of work is the idea that institutions (voting rules, types of participation, distribution of speaking rights, nonpolicy side payments, etc.) have consequences for individual choices and thus for social outcomes. Because these outcomes are likely to differ with respect to their epistemic and welfare properties, a normative theory of deliberative democracy must be a theory of institutions. A failure to recognize this point is responsible for the view that the normative arguments of deliberative democrats are somehow inconsistent with the notion of individual rationality. According to this view, most recently advanced by Posner (2004), if individuals are choosing the nature of their deliberative engagement rationally, then their choices imply that any further argument in favor of deliberation must, in effect, counsel behavior that is contrary to rationality.

It is important to see that this view rests on a fundamental mistake about the determinants of social outcomes. Different institutions provide individuals with different incentives by, *inter alia*, affecting the size and the nature of one’s potential audience (such as electoral campaign subsidies, time in front of the microphone in a committee meeting, etc.), changing which voter is pivotal (by changing the voting rules, the degree of centralization in collective decision making, etc.), distributing decision-making authority across levels of government, etc. Although the incentives these institutional choices create are not always immediate to outside observers, individuals who are subject to them can be expected to recognize them over time and adapt their choices regarding deliberation accordingly. To evaluate institutions with respect to properties of deliberative outcomes is to ask whether they create incentives such that, in the aggregate, the outcomes they give rise to are as desirable as those attained under other institutions.

This argument underlies both a causal claim about institutions and a methodological recommendation in connection with normative arguments regarding deliberation. The former is by now clear: institutions affect behavioral choices and, through them, the properties of deliberative outcomes. The latter is, arguably, implicit in the causal claim: normative arguments regarding institutional choice must treat deliberative behavior as, *inter alia*, dependent on the institution. If they do not, there is no reason to believe that the deliberative and policy outcomes will, in fact, have properties that are consistent with the expectations associated with the endorsed institutions.

¹¹Our view, nonetheless, contrasts with how empirical scholars who eschew game-theoretic work see the relationship between normative theory and the empirical analysis. These scholars seek to evaluate normative accounts of the effects of deliberation based on whether observational or experimental data are consistent with those accounts (Mendelberg 2002; Mutz 2008). While the empirical evaluation of theories is surely an essential part of a social science of deliberation, we see game-theoretic analysis as an indispensable element of that evaluation. As Mutz (2008) argues in criticizing the unfalsifiability of normative theories of deliberation, to test a theory of deliberation, one must first assemble a coherent set of antecedents and identify their status as necessary and/or sufficient conditions for a communication to have particular posited properties. But that is exactly what the analysis of axiomatic consistency delivers: by using what we called strategic behavioral consistency as a guide, game theorists construct conditions of coherence and incoherence for different normative accounts.

¹²Work on this question is nearly nonexistent (but see Knight and Johnson 2007; Landa 2005 List n.d.).

Incentives in Deliberative Environments

Discourse, Policymaking, and Incentives

In comparing deliberative environments, an immediate question to confront is whether the focus of analysis is on discussion and debate or on discussion and debate by participants who are making policy decisions (sometimes referred to as decision making “in committees”). As we discuss in greater detail below, recent findings by game theorists suggest that this distinction has critical consequences for what may be expected from the deliberative process. When communication precedes policymaking, participants will have incentives to misrepresent or withhold information unless their underlying values or preferences are commonly known to be quite similar. In contrast, if one considers a deliberative environment in which discussion is not followed by policymaking, and participants care only about whether they themselves arrive at the most defensible judgments, it can be easier to sustain informative, truthful debate. This difference in incentives is important not simply because game theorists care about strategic behavior, but because, as we emphasized in the preceding discussion, the institutions of policymaking may be expected to have consequences for individual choices, including individual choice in the debate prior to decision making. Whether the debate is followed by policymaking and who is making the decisions become critical factors in ascertaining what behavior may be expected in the debate.¹³

Even if there is a distance (and that distance may well, and in important ways, be consequential behaviorally), deliberation in politics is, arguably, best understood as ultimately leading to policy choices. This is not to say that traditionally idealized deliberative forums like New England town-hall meetings are unimportant aspects of democracies. Indeed, they are important, but understanding their role turns on two issues that are inextricably tied to decision making. The first issue—a mix of substantive and methodological concerns—has to do with the informational effects of deliberation in such forums. To the extent that deliberative democracy

is representative democracy, policymaking in a committee setting will remain one of its key concerns. To make sense, for example, of the effects of town-hall meetings on such policymaking, we have to rely on the account of how information is shared and aggregated in committees, including the ability to isolate factors that determine *how much* and *whether* it is—factors that may be tied to the policy effects of town-hall meetings. Without knowing the counterfactual (what the deliberation would be like in the absence of those factors) and understanding the channels of influence, we cannot know what the real effects of deliberation in town-hall meetings are.

The second issue concerns the plausibility of interpreting deliberation in such a forum as a case of “pure discourse.” Indeed, the nature of incentives faced by individual participants may be more consistent with what would generally be construed as “committee” deliberation: participants aim to promote the beliefs and arguments they consider compelling because they would like to affect the positions of other decision makers; to the extent that deliberation is consequential for what policies are chosen, those positions, whether ultimately regarding collective choices (e.g., whether to protect the right to obtain abortions) or individual choices (e.g., whether to obtain an abortion), have externalities—that is, they affect how participants perceive their welfare and create incentives for them to influence the selection of more favorable alternatives. In this view, town-hall meetings are instances of deliberation leading to policy choice rather than of “pure discourse.” If our expectation of what deliberation in town-hall meetings is like suggests a relatively unproblematic communication, then what makes this so is an important puzzle, and providing a systematic explanatory account of it should be one of the important tasks of a theory of deliberation. (As our discussion below indicates, precisely what factors are responsible for it is not altogether obvious.)

With these understandings in mind, in the remainder of this section we focus on the distinct environments in which participants debate and ultimately select a policy, and consider some of the diversity of incentives that can surface in them.¹⁴

¹³Because formal models require an explicit description of the environment, it is clear that positive theorists have been almost exclusively (Glazer and Rubinstein 2001 is an exception) interested in debate that precedes policymaking. In normative and social psychology works, the distinction between discourse and deliberative democratic policymaking is not always clear. While some scholars theorize deliberation as an institutionalized part of an explicit electoral process, they seek supporting evidence in the analysis of deliberative practices in contexts in which the relationship to policymaking decisions is either absent or ambiguous (e.g., Barabas 2004; Fishkin 1991).

¹⁴To the extent that this is our focus, it is, of course, worth asking what evidence there is that deliberation is at all policy consequential. That it is consequential, at least under the specified conditions, is unproblematic in the theoretical analysis (whether game-theoretic or normative). Given the difficulty of establishing a well-defined “control” or counterfactual in actual policy settings, it is considerably harder to show this empirically. Still, a number of recent studies point to evidence across policy areas and institutions suggestive of the presence of distinct policy effects. Thus, Epstein and Knight (1998) and Lax and Rader (2007) provide evidence that members of the U.S. Supreme Court switch their allegiance to

Common Values, Noncommon Values, and “Cheap Talk”

Consider a parlor game that might be a small step closer to policymaking than poker is. A group of individuals must decide as a group whether to bet on a particular die coming up even or odd. Each group member gets a share of the group’s winnings. The individuals in the group are uncertain of the odds of the die coming up odd or even (i.e., how fair or unfair it is). Prior to betting each member gets to privately observe one toss of the die. The group members then assemble and are given the opportunity to talk and vote over which way the group should bet. Assume that the group members all want to maximize their earnings and that all members of the group know this. In this environment we would expect the group discussion to be much more informative than in a poker game. In fact, we might expect them to be as informative as possible, truthfully revealing what they observed and discussing which bet is best. It is possible to show that these expectations are consistent with the formal requirements of equilibrium play.¹⁵

In reaching this conclusion, we made a critical assumption: it is commonly known that all members of the group get a share of the group’s earnings and so are interested in maximizing it. To see why such specificity is important, suppose that one of the participants, say player 1, receives a positive payoff if the group’s bet is incorrect and a negative payoff if the group’s bet is correct. In this case player 1 might have an incentive to misreport the outcome of the toss that she alone observed, in hopes of leading the group to the wrong decision. In a setting in which the other players do not know that player 1 has this conflict of interest, player 1’s report might be believed. If, however, all of the players know that player 1 has a different motivation, then there cannot be an equilibrium in which she is taken at her word. If player 1 could say something that would lead the other participants to believe that a particular outcome were more likely, then she would want to send this type of message whenever she

opinions (sometimes, reversing the holdings in the process) in response to specific arguments advanced by the opinion writers. Sunstein, Schkade, and Ellman (2003) find evidence that, holding fixed the ideology of the majority, ideological composition of three-judge panels on courts of appeals (interpreted by the authors as implying an informational effect in internal deliberations) affects panels’ decisions across issue areas. Meade and Stasavage (2008) link the recently implemented public release of meeting transcripts of the Federal Reserve’s Federal Open Market Committee with the change in the quality of the deliberation on it and with the effects of monetary policy institutions more broadly.

¹⁵Other types of behavior, including speeches that are entirely uninformative, are also consistent with reasonable equilibrium concepts.

thought that outcome was unlikely. Depending on the circumstances, these incentives may be thought to resemble those in our example of the debate before Iraq invasion. If the source of the key relevant classified information is the members of the prowar camp who also happen to be strongly committed to using military force to spread democracy, then statements from them that a good evidentiary case for the presence of WMDs exists should be considered to be similarly uninformative.¹⁶ Plausible interpretations of reality in light of statements made by some speaker will, thus, depend on what that speaker’s preferences are thought to be; it matters not just what preferences one has but also what beliefs other players have about one’s preferences.

A group in which it is known that all participants have common preferences and a group in which it is known that participants have opposing preferences represents the opposite ends of a spectrum. Game theorists typically use the term *common values* to describe the former case and *noncommon values* the remaining cases, and distinguish between the incentive problems presented by each. The good news is that in common values problems, honesty is a reasonable expectation. More precisely, there are equilibria in which participants truthfully reveal what they know and the voting reflects all of the available information. The bad news, however, is that in noncommon values problems, expectations of honesty may be unwarranted (Meirowitz 2006, 2007).

Why should a deliberative democrat be concerned with the distinction between noncommon and common values? One of the central issues of contention among deliberative democrats is the expectation of consensus following deliberation. Jürgen Habermas (e.g., 1996) has argued for the reasonableness of such an expectation—in particular, that, if participants could argue indefinitely, they would converge on the same judgments. The preceding discussion suggests that given rational agency, that expectation is either wrong or its validity hinges on the assumption of common values. It is important to emphasize that common values does not mean merely identical primitive preferences. It is far more demanding: it requires that the interlocutors have common knowledge of this identity (that is, that each of them knows that the others know that he knows that primitive preferences are

¹⁶To be sure, it may not be entirely accurate to call that “lying.” These agents may have less of an incentive, on the margin, to invest into getting accurate information. Consequently, the information they do have could very well amount to signals supporting their claim, yet it is also likely to be both noisy and biased. Because confirmatory biases may systematically skew their personal assessments of the evidence, they may well not be cognizant of these properties of their data.

identical). Thus, the very fact that the scholars of deliberative democracy disagree on this point (see, e.g., Bohman and Rehg 1996) strongly suggests that the assumption of common values is untenable (even if participants *actually* have the same preferences). To be slightly tongue-in-cheek, we can say that any collection of individuals that includes the published theorists of deliberation does not have common knowledge of common values.

If the paradigmatic deliberative interaction involves noncommon values, then the above discussion suggests the importance of taking seriously the incentives faced by the participants with respect to their deliberative choices: what, if any, information and arguments to share with others, whether to engage in misrepresentation, and how to interpret the communications of others. This underscores the importance of two questions about deliberative environments that are central to game-theoretic models: When is it the case that there are equilibria in which participants are truthful in their speech making? What are the properties of the equilibrium policy choices? Once these questions are settled, the deliberative theorist can go on to determine, given the adopted set of normative criteria, the best type of environment to implement for deliberation.

Incentives in Richer Deliberative Environments

Both in the poker and in the odd-even die examples, speaking imposes no costs on the speaker, and that means that the listener or observer cannot infer which speeches are truthful and which are not from the speaker's cost of making them. Such "cheap-talk" statements are a frequent feature of social and political interactions in democratic institutions.¹⁷ The problems of (in-)sincere speech, and consequently, of underinformed postdeliberative decision making, that arise in environments without common values may be commonplace in these settings.

Still, cheap-talk arguments are not the only kinds of arguments that could be made in deliberation, and this naturally raises the question of the scope of deliberative interactions with the strategic problems that we described above. This question has a further philosophical dimension identified by Johnson (1993). As he notes, problems of indeterminacy that sometimes arise in game-theoretic contexts—mainly as a consequence of strategic

uncertainty in environments with multiple equilibria—suggest that strategic action may be an insufficient source of social coordination. While allowing the possibility of cheap-talk communication can lead to the selection of more efficient equilibria, it does not do so in a social vacuum. Effective cheap-talk communication presupposes prior common comprehensibility of signals (a notion of common language) that are also consistent with equilibrium play. This suggests the presence of limits on cheap talk as *the* model of verbal communication.¹⁸

Some of the recent work on game-theoretic models of deliberation is, in part, motivated by the same concern and, arguably, takes the road Johnson recommends—focusing on the binding force of argumentation. In cheap-talk deliberation, the mechanism for inference relies on the listener's ability to pin down the determinants of utterances. That is, the listener must answer the following question: given what is known about the nature of the deliberative interaction, including the preferences and the beliefs of the participants, what kind of private information could the speaker have that would be consistent with the observed speech, given that the speaker is making rational decisions about what to say (that is, understanding how her speech will likely be perceived)? In order for an equilibrium to involve learning, the listener(s) must correctly believe that the speaker(s) would like the listener(s) to learn what is learned, and that the messages sent have particular meanings commonly recognized as such by the speaker and the listener. The underlying idea behind moving away from cheap talk is that if argumentative support could be brought to bear in communication, then a range of messages that may be sent credibly would increase and so the actual speech may become more informative. The mechanism for inference in such speech differs substantially from the standard cheap-talk environment: the listener may determine the validity of the speaker's statements directly, through the consideration of undeniable evidence.

Arguments in deliberation may display different positive degrees of direct (or intrinsic) provability. An argument could be *partially provable* in the sense that only a part of it is verifiable, leaving some residual

¹⁷This usage, which has become standard, combines two logically distinct attributes: costless and nonverifiable communication. Below we also consider costless but verifiable communication. Our reference to it as an alternative to "cheap" talk is in keeping with the standard usage.

¹⁸As Johnson points out, in "real-life" natural-language interactions in environments with noncommon values, speakers may have both incentives and the ability to subvert the equilibrium play by using messages that are not anticipated in equilibrium—if they can get the listeners to interpret them in ways that are more beneficial to the speakers than the equilibrium interpretation of (expected) messages. The consequence is that, while speech is often able to coordinate expectations effectively, the standard cheap-talk model of communication may be missing elements that "might compellingly account for the binding force of language in strategic interaction" (1993, 82).

uncertainty about its validity. The argument could also be *fully provable*—that is, we could know the full merits of the argument (though it may also happen that we could obtain this proof of validity with some positive probability that is short of certainty). The provability of arguments may stem from their logical persuasiveness or unfalsifiable hard evidence. The possibility of argument provability may at first suggest that the concerns about strategic dissembling that are rampant in the cheap-talk environment with noncommon values may be irrelevant. Alas, in connection with partially and fully provable arguments, noncommon values raise another set of strategic concerns. Verifying the validity of arguments is typically costly—it requires an investment into becoming informed (either literally paying for corroboration, or incurring the direct cost of acquiring expertise on one's own, or incurring the opportunity cost of listening and processing when a claim is being justified with a potentially provable argument). The existence of such costs means that a speaker with a primitive preference that is potentially different from that of her audience will sometimes want to make arguments that are provably wrong in the hope of passing them off as (possibly) right ones. In such cases, the speaker may rationally anticipate that the listener will not incur the cost of verifying the argument, taking it as true with some positive probability. Alternatively, if verifying the argument would require sacrificing resources that may otherwise be spent in ways contrary to the speaker's preference, the speaker may have a further interest in making provably wrong arguments in the hope that the listener incurs such a cost (Landa 2005).¹⁹

More striking, however, is the fact that even when arguments are instantaneously and costlessly verifiable, incentive problems surface. The following example provides an illustration. Suppose a group of committee members are deciding whether to support Jill, a candidate for office. A majority of the committee members are concerned about Jill's integrity but like her policy ideas. They will support her unless they have clear evidence of improprieties. Jack, a well-connected member of the committee, happens to know Jill well, and thus might have some additional information about Jill. It is possible that he is aware of indiscretions and can provide documentation to the committee. It is also possible that Jack possesses no additional information about Jill's past. Suppose that Jack is exceptional in his tendency to not let a person's past in-

fluence his assessment of his or her ability to perform in the future; Jack believes that regardless of her past Jill is the best person for the job. What should we expect Jack to do if he actually does have evidence of Jill's past indiscretions? In this noncheap-talk setting, Jack will strategically refrain from providing information that is valuable to the other committee members if he places enough weight on his belief that Jill is best for the job and little weight on being forthcoming (or, if his fellow committee members' judgments could affect his own, on the alignment of his primitive preferences and theirs). Knowing this, how will the other members interpret *silence* by Jack? If, regardless of whether Jill has, in fact, been guilty, Jack would remain silent, then his silence cannot be taken as evidence that he does not know of any indiscretions by Jill, and thus the committee loses the ability to use Jack's expertise about Jill to its fullest extent. Driving the incentive problem is the fact that, even when arguments are provable, the speaker has a choice about whether to make the argument. There is no slight of hand here; a speaker can refrain from providing a provable argument about a question as long as there is uncertainty about whether she knows the answer.

Parallel to the distinction between common and noncommon values, there may exist one with respect to the perceived truth content or persuasiveness of arguments. Say that an environment is characterized by *common veridicality* if it is known that all participants would agree as to which arguments are persuasive and which are not; *noncommon veridicality*, then, corresponds to the case in which such agreement may not exist. Although we tend to associate positive provability with common veridicality (e.g., if Fermat's last theorem is true, it is true for everyone), deliberations in politics often combine provability with noncommon veridicality. The intuition for this possibility is that if citizens' moral commitments or values are sufficiently different, this difference may be expected to affect what arguments they find persuasive (e.g., for and against legalization of gay marriage, for and against racial profiling, etc.). In such cases, arguments that combine provability with noncommon veridicality often proceed by articulating logical consequences of claims that some part, but typically not the whole, of the audience accepts as true but, for whatever reason, mistakenly considered irrelevant to the issue at hand. The familiar expression "good-faith disagreements" may be thought to refer to the case of common values paired with noncommon veridicality. In contrast, debate between experts with vested interests may be thought to involve the opposite pairing, of noncommon values and common veridicality.

In the environment with common values and common veridicality, deliberation may be seen as a purely cooperative endeavor: setting aside the costs of deliberative

¹⁹In the case of noncommon values and costly monitoring, if, in fact, equilibria with full revelation exist, then monitoring must occur with positive probability. This monitoring itself represents a form of inefficiency—a point missed by the claim that monitoring solves the incentive problems present in cheap-talk models (see, most notably, Mackie 1998).

engagement, individuals may be expected to seek to convince their interlocutors by sharing with them everything they know (Coughlan 2000). Noncommon veridicality changes this incentive. If deliberation is followed by policy choice, then even in the cases of would-be good-faith disagreements, deliberators have incentives to avoid making a (dis-)provable argument if doing so may convince the listener that she is the sort of agent who finds that sort of argument unpersuasive and that she should make a policy choice that reflects that. Even when agents have aligned preferences, noncommon veridicality may result in incentives to withhold information.

The following example provides an illustration. Suppose that Jill is considering getting an abortion. She leans against it on moral grounds, but is not yet convinced that she has heard the most persuasive arguments that would support that position. She also recognizes that she may not have heard the most persuasive arguments against it. Suppose, further, that Jill finds herself in a discussion with Jack, who she has good reason to believe is one of the most thoughtful critics of abortion, and Jack makes to her a series of arguments, none of which she finds persuasive. How should this interaction affect Jill's beliefs about the justifiability of her prior position? She should conclude that the decision to abort may be more defensible than she had previously thought; here is one of the most compelling critics and his arguments are not particularly persuasive. Knowing this, and knowing that Jill would otherwise be likely to go through with the pregnancy, Jack may be reluctant to make his arguments to her. In short, in the case of noncommon veridicality, the speaker will often have an incentive to avoid making provable arguments (Hafer and Landa 2006), and in equilibrium, astute listeners will discount the arguments they hear, thinking them uninformative (that is, speaking to neither the truth nor the falsity of the relevant claims). The incentive to lie in the "cheap talk" setting has a counterpart as an incentive to refrain from providing arguments in settings with provability.

The preceding discussion suggests that the kinds of strategic incentive problems that arise in the cheap-talk environment arise in other informational environments as well. Because criteria of deliberative performance in democratic societies must surely include the extent to which the postdeliberative judgments of citizens are educated by the deliberative process—that is, the extent to which they are close to what they would be if such a process resulted in the articulation of all the relevant information and arguments—the presence of incentives to refrain from making provable and fully informative non-provable arguments underscores the importance of equi-

librium analysis for normative theories of institutional choice.

Are Game-Theoretic Conclusions an Artifact?

The recent game-theoretic work on deliberation advances the formal analysis of communication by focusing on the effects of different informational structures and incentives arising when multiple interested parties possess information that is relevant to the collective choice. The possibility that noncommon values might lead to incentives to mislead others has, however, been a prominent fixture in the economics literature since the seminal work by Crawford and Sobel (1982) and the early applications to the politics of debate (Austen-Smith 1990; Austen-Smith and Riker 1987). Despite the appearance of a number of formal studies exploring related ideas in the intervening time, normative theorists have, as we noted above, largely dismissed these concerns—typically by simply ignoring the existence of the parallel research tradition, but, in exceptional cases, including Heath (2001), Cohen (1998), and Mackie (1998), by giving an explicit reasoned critique. Our analysis of this critique identifies six arguments—three concerning particular aspects of the deliberative environment, and another three concerning the nature of deliberative behavior, which we take up and address in the remainder of the article.²⁰

Strategic Deliberative Environments

Before considering the particular arguments, we want to highlight a key general feature of our response. The three arguments about the environment that we address have the following common flavor. They begin by observing that the canonical or extant models have oversimplified reality; each points to elements of some actual deliberative settings that are absent from the game-theoretic models. They proceed, then, to argue that, in a richer model, incentive problems may be ameliorated. Our central objection to these arguments is that, at best, they are correct that in some models with these additional elements incentive problems diminish. They do not, however, demonstrate that in the reasonable subset of models with these added elements, incentive problems are diminished. Indeed, we

²⁰We set aside objections to the theory of rational choice as such, a discussion which is outside the scope of the present article (though see the discussion of behavioral agency above).

provide reasons for thinking that it may well be otherwise. However, criticisms of exactly this form (such and such a model is not to be taken as the correct representation because it ignores something that is important) are fundamental for developing better theories of deliberation. The fact that these criticisms are taken as grounds for ignoring game theory, instead of grounds for improving its application to the study of deliberation, is where scholarship on deliberative democracy breaks down.

The first critical argument, articulated by Mackie (1998, 84–85), proceeds by observing that participants care about their welfare tomorrow as well as today. While deception today might result in a more desirable policy outcome, it will lead to a reputational hit or other future punishment. The problems of information revelation raised in game-theoretic models are thus, to a considerable extent, an artifact of failing to incorporate the long-term nature of the interaction.

While sometimes plausible, this conclusion runs into problems at other times. The first problem is that, while it is true that repeated play may enlarge the set of behaviors that are consistent with equilibria (a paraphrase of the folk theorems), it is typically the case that equilibria with dishonesty survive. Thus repetition may introduce the possibility of better information aggregation, but it cannot rule out the possibility of poor aggregation. (As an analog, note that repeated play of the Prisoners' Dilemma does not eradicate the "bad" equilibria, it just introduces the possibility that there are also "good" equilibria.)

A far more striking problem is that, although repetition makes reputational concerns in principle possible, it does not necessarily make sanctioning a participant who behaves poorly feasible. Sanctioning is effective only if participants can determine whether an individual has revealed her private information. Our first Jack and Jill example above illustrates this point. If only Jack knows whether he has an argument to make, punishing him if he does not make one will not improve the informational quality of deliberation. In fact, in such circumstances, game-theoretic analysis of the effects of repetition suggests that repetition can, in fact, make things worse, not better. Morris (2001) considers a policy maker seeking advice from an advisor on a policy decision and shows that the shadow of the future can actually create incentives for dishonesty. The following motivating example captures the intuition:

Consider the plight of an informed social scientist advising an uninformed policy maker on the merits of affirmative action by race. If the social scientist were racist, she would oppose affirmative action. In fact she is not racist, but

she has come to the conclusion that affirmative action is an ill-conceived policy to address racism. The policy maker is not racist, but since he believes that there is a high probability that the social scientist is not racist, he would take an anti-affirmative action recommendation seriously and adjust government policy accordingly. But an anti-affirmative action recommendation would increase the probability that the policy maker believes the social scientist to be racist. If the social scientist is sufficiently concerned about being perceived to be racist, she will have an incentive to lie and recommend affirmative action. But this being the case, she would not be believed even if she sincerely believed in affirmative action and recommended it. Either way, the social scientist's socially valuable information is lost. (231–32)

The logic behind this example is particularly damaging to the reputational argument because it is precisely the advisor's concern for her reputation in the eyes of the policy maker that furnishes her with an incentive to lie, even though both players, in fact, have the same preferences. In sharp contrast to the case in which participants can observe when a speaker has been dishonest and punish her, if such observations are not possible, reputational concerns can make matters worse. To be clear, our conclusion here is not that the shadow of the future never improves the quality of deliberation—indeed, it can—but rather that how, when, and at what cost is not at all obvious without further analysis. Turning even a plausible intuition, like the one about the positive effects of repetition, into an argument requires precisely the kind of detailed analysis of incentives that game theory presents.

The second argument against the relevance of game-theoretic studies of deliberation is that these studies restrict their attention to a single speaker. With multiple speakers, Mackie (1998) argues, incentive problems are moot. In particular, he argues that when there is a "trusted" speaker (somebody whose preferences are similar to the receiver's and this is known to both the senders and the receiver), a nontrusted speaker cannot deceive a rational listener. Mackie's argument, however, hinges on the assumption that the receiver knows a priori which agent is "trusted," i.e., which speaker shares the receiver's preferences. In fact, as long as preferences are common knowledge, having two senders with opposing (but not necessarily symmetric) preferences can be enough to generate informative speech *regardless* of whether there is a sender who shares the receiver's preferences (Battaglini 2000; Krishna and Morgan 2001). Alas,

some of the thorniest long-standing problems in politics arise precisely because citizens do *not* know the preferences of others, and thus whom to trust. Indeed, a central justification for political science's focus on the analysis of institutions, including deliberative mechanisms, is that institutions represent a means to solve such problems. A defense of deliberation that assumes that participants necessarily know each other's preferences seems to put the cart before the horse.

An important finding of recent work on communication is that when information about the true preferences of the senders is incomplete, then information transmission is no longer possible in equilibrium. Minozzi (2006) analyzes a game with two senders and such an information structure and finds that speakers with opposing interests have incentives to send messages that contradict each other in a way that makes it impossible for the receiver to infer the truth even when one of the senders is actually truthful. Other game-theoretic analyses of deliberation with multiple speakers reinforce the mixed-record status of expectations of informative equilibria. Meirowitz (2006, 2007) shows that under some strong conditions, having multiple participants can help. In particular, if everything that is known by one participant is known by at least two other participants and this is common knowledge, then there are equilibria that fully aggregate the information—as well as equilibria that do not aggregate the information. However, these “nice” results are restricted to special circumstances, and Meirowitz (2007) and Dickson, Hafer, and Landa (2008b) demonstrate that when the speaker faces uncertainty about the preferences of other speakers, then having more participants may, in some environments, not only fail to help, but, in fact, be detrimental for eliciting informative speech.²¹ The special circumstance in Mackie's example is a move from the case of one expert with preferences that are not aligned with the receivers to a case with an additional agent possessing exactly the same information but also sharing the receiver's preferences. Such a move will indeed make the receiver better off, but many other approaches to changing the number of experts will lead to quite different conclusions. In the end, the multiplicity of participants is neither necessary nor sufficient for ensuring informative communication.

The third argument holds that a fundamental shortcoming of the formal analyses of deliberation is their re-

liance on the “cheap talk” technology; with reasoned (or provable) arguments, distortions associated with strategic action with respect to deliberation would be mitigated if not moot (Cohen 1998; Mackie 1998).²² A formal support for this view is based on Lipman and Seppi's (1995) model in which participants can make partially provable arguments. Their equilibrium construction hinges on showing that if one speaker lies, another speaker with different preferences might come along and show that she has lied and then influence the outcome. In response to this fear, the first speaker is better off being honest straight away. This result hinges on the assumption that all the speakers possess the same information. In particular, Lipman and Seppi do not show that the presence of multiple speakers and partial provability will create incentives for agents to share information that is not also possessed by others in the group. Indeed, as the Jack and Jill examples of the previous section demonstrate, even full provability is not sufficient for eradicating incentive problems in information revelation and argument making. When their arguments are provable, participants will have incentives not to share their information if they anticipate that the deliberative environment is one of noncommon veridicality. What matters is the existence of some form of underlying disagreement—whether directly interest based or epistemic. When such disagreement is present, the incentive problems may be expected to persist.

In short, simply adding features like multiple senders, repeated play, or provability to the canonical strategic models of communication does not justify the conclusion that deliberation in environments with these features works well. Other features of the deliberative environments—features that vary across empirical political circumstances—should fundamentally affect our expectations about deliberation. However they may do so, it is also important to see that just because incentive problems exist in deliberative settings, it does not follow that deliberative institutions are undesirable. After recognizing that no perfect institutions exist, we are left with the challenge of determining which imperfect institution is best. More broadly, to focus on the complexities of strategic incentives to reveal information is *not to argue against deliberation*. Rather, as we noted above, elucidating the features that influence the informational properties of deliberative behavior is instrumental to the effective design of good (or at least the best feasible) institutions. At the

²¹Meirowitz shows that with sufficiently few participants, all speakers will be honest, but as the number of participants goes up, certain types of speakers will sometimes prefer to dissemble. Dickson, Hafer, and Landa find that under particular rules of debate and voting agendas, the prospect of counterarguments from ideological opponents discourages argumentation.

²²Note that this line of critique is distinct from that of Johnson (1993) that we considered above. We take the upshot of Johnson's point to be that binding speech acts resolve particular problems for the theory of strategic communication, rather than that the possibility of binding speech acts eliminates the possibility of strategic action with respect to communication.

end of the day, it makes it possible to know which features of deliberative institutions to choose in order to make the most of deliberation. In this sense, by observing that particulars of the model drive inferences about behavior, critics of the game-theoretic approach are encouraging precisely the institutional analysis of deliberative democracy that we are urging in this article.

Strategic Deliberative Behavior

It is, perhaps, unsurprising that the critiques of the game-theoretic conclusions that focus on strategic *deliberative behavior* are somewhat more philosophic in nature. As a class, these critiques argue that the behavior characterized in the strategic models is inconsistent with the principal sources of participants' motivations, which tend to fall outside those models, and that such motivations trump the motivations explored by game theorists.

One such argument anticipates that the process of deliberation may lead participants to set aside their self-interested motives in favor of reasoning with respect to a parallel set of desiderata—those that are more appropriate to the deliberative interaction. “Seeing that certain of my antecedent preferences and interests cannot be expressed in the form of acceptable reasons may help to limit the force of such preferences as political motives” (Cohen 1998, 199). This may happen for two reasons: either because we should be normatively committed to the view that “political justification requires finding reasons acceptable to others, understood as free and equal, who endorse that commitment” (Cohen 1998, 200; see also Scanlon 1998, chap. 5), or because of what Elster (1995) calls “the civilizing force of hypocrisy”—we may simply be forced into making less self-serving arguments to argue successfully with others and over time may internalize the value of other-regarding reason-giving as a way of coping with dissonance.

Elster's hypothesis would, likely, find some psychological support. However, there are good reasons to resist building a normative theory of deliberation upon this premise: it is equally compatible with conformity and raises questions about the compatibility of democratic procedures and individual autonomy (Johnson 1998, 172). A normative commitment to “honest” or “fully revealing” deliberation does not suffer from these problems, but in a complex environment with private moral values and noncommon veridicality, it must presuppose not simply abnegation of self-interest, but an extraordinarily high degree of abnegation of moral instrumental behavior as well. In both of our Jack and Jill examples, Jack's motivation is not self-interested: he is dedicated to a particular set of values and policies because he believes

them to be best on the merits. To the extent that, mindful of noncommon values or noncommon veridicality, Jack believes that deliberation will not lead to a consensus, in which either he and/or his interlocutors will necessarily be convinced by the other, a requirement that he always fully and honestly reveal may be expected to run into conflict with his commitment to the values and policies that he believes are most defensible. It is that extremely demanding and quite controversial requirement that is, in effect, urged by this objection to the game-theoretic arguments. Embracing this requirement as a default seems premature both with respect to the status of moral-theoretic debate on its validity and as an empirical description of individual motivations.

Another argument targeting the deliberative behavior characterized in game-theoretic models concerns the legitimacy of the postdeliberative outcome. Its claim is that we want such outcomes to be perceived as legitimate and that requires that participants fully and honestly reveal the arguments and information available to them (Habermas 1990). This claim is, in effect, based on a comparison of two states of the world: in state 1, the speaker, say, Jack, fails to speak or fails to reveal all the information available to him, and the outcome is less legitimate; in state 2, Jack reveals more information, and the outcome is more legitimate. To make sense of this argument, we must suppose, further, that, save for the effect on political legitimacy, Jack would prefer state 1 to state 2; if it is not the case, then the invocation of political legitimacy is moot. It should be clear now that, in order for the argument to have bite, it must effectively endorse the following two claims: (1) interest in the perception of the outcome as legitimate overrides whatever interests Jack may have in revealing less information; and (2) perception of the outcome as legitimate is responsive to how much information Jack reveals. Both of these claims strike us as highly controversial and contingent, at best. The first claim runs into a problem similar to the one we raised in connection with the previous argument: it presupposes a considerable constraint on Jack's commitment to the values or policies he believes to be most defensible. The second claim assumes, in effect, that it is known when Jack has relevant information and refuses to share it—an unlikely event in the context of our first Jack and Jill example, and certainly unreasonable as a default supposition.²³

²³There is another variation on this argument (see Fearon 1998, 62, for a brief discussion): we want political outcomes to be perceived as legitimate, and legitimacy is enhanced by having losers in a vote “know exactly what reasons and arguments the winners had judged to be stronger in deciding the merits of the case.” Here, there are two implied counterfactual states as well. In state 1, information is

The last critical argument we consider proceeds on the premise that, in the same way that telling a lie is parasitic on telling the truth, strategic action is parasitic on “communicative action”—action that is oriented toward reaching mutual understanding in the course of deliberation—in the sense that the possibility of strategic action requires communicative action (Habermas 1984). Thus, the incentive-based behavior that game theorists are focusing on is, in essence, of secondary importance; it cannot, *ipso facto*, be the core of the deliberative practice.

In part, this argument echoes Johnson’s challenge, discussed above, to account for binding communication that belies the cheap-talk model. As we argued in the preceding discussion, deliberation with partially or fully provable messages meets this challenge but does not eliminate incentive problems in argumentation. Heath (2001, chap. 2), however, pushes the argument that strategic speech is (relatively) unimportant in a different direction. He succeeds in making the case that when informative speech is ruled out in equilibrium, game-theoretic analysis of communication cannot successfully explain information transmission if or when it should occur. As he notes, this is in part because in a game-theoretic model of communication, agents’ goal is not “idle speech” merely aiming to affect beliefs of other agents but, rather, speech aiming to affect particular outcomes. Because proximity of preferences over outcomes (i.e., whether the agents have sufficiently common values) determines agents’ interests in information revelation, sufficiently noncommon values lead to uninformative communication in equilibrium. Insofar as communication is, in fact, informative, this outcome would indicate that the strategic model of communication (cheap talk as the presumptive target) is lacking in some fundamental way.

At the end of the day, how much of a limitation is entailed in the possibility of failing to account for informative speech in such cases seems to us to turn, in part, on one’s fundamental account of politics. If politics is fundamentally noncontentious (strictly common values), strategic speech is often simply moot; if it is supremely contentious (diametrically opposed preferences), a strategic model of speech is unhelpful, and if social coordination occurs via verbal communication, it must do so in ways that are best analyzed with other analytical tools.

Most everyday politics is, however, more likely than not somewhere in between—preferences do not exhibit

not revealed, Jack is better off, but losers do not know what it is that they lost to; in state 2, information is revealed, Jack is worse off, but losers know what it is that they lost to. To the extent that Jack has something to hide, it seems to us that the outcome in state 2 would be perceived as less, not more, legitimate.

common values but there are degrees of preference alignment present (Hardin 1988). In that middle range, politically relevant communication is rarely without a purpose of affecting an outcome; and its informativeness is, from the standpoint of game-theoretic models, both possible and not random. As a theory of rational action, game theory is committed to the view that that informativeness relates systematically to the incentives faced by the participants (even when sincerity does not come about as an intentional choice, it may be a consequence of the fact that the underlying environment is such that the incentives to misrepresent are not prominent). In effect, the impulse behind the game-theoretic analysis of deliberation is to “earn” the sincerity by reconstructing it as equilibrium behavior rather than assuming it by default. As we argued above, the value of doing so is not only explanatory. Unless we understand the conditions under which the incentives in deliberative environments encourage agents to be sincere or fully revealing, as opposed to insincere or withholding information, we cannot hope to offer a coherent (stable) normative argument for institutional design. If one of the goals of normative theories of deliberation is to offer such arguments, the game-theoretic analysis of incentives in deliberative environments is a useful tool that should be of primary importance even if strategic action is parasitic in the above sense.

Conclusion

Game-theoretic analyses demonstrate that deliberative democracy is not immune from incentive problems. Specifically, incentives to misrepresent or to refrain from sharing relevant information surface in many analytically distinct types of settings. What consequences should these observations be taken to have for the further development of deliberative democratic theory?

One possibility is to ignore the problems and to study only sufficiently idealized deliberation. To the extent that expectations of sincere speech are critical for the deliberation-based theory of political legitimacy, this escape seems self-defeating, limiting the applicability of normative claims about deliberation to the most trivial of collective choice problems.

A second option is to accept the possibility of incentive problems but proceed on the assertion that these problems are only consequential in a narrow range of settings. As we have emphasized, however, incentive problems surface even where critics argue such problems have no bite—noncheap-talk settings, settings with repeat play, and settings with multiple senders and/or receivers. This option, then, also seems problematic.

A third response is to concede that these problems are important, caveat one's arguments and conclusions with the assumption that incentive problems have been or will be addressed and continue the research agenda without modification. This course of action is, in our opinion, ill advised. The study of incentives leads to insights about how the particulars of a setting influence behavior, and, at present, these particulars tend to be absent from normative accounts. Unless we are convinced that these details are irrelevant, the current literature is likely to suffer from the pathology of formulating one-size-fits-all prescriptions. Based on the extant game-theoretic treatments, there is reason to believe that the details can matter a great deal, and ignoring the study of incentives means missing an opportunity to develop richer, more nuanced theories of deliberative democracy.

We have sought to make the case for the value of a fourth option—one that goes beyond the friendliest of the “live and let live” responses. It draws on the normative analysis to elucidate axioms and rank trade-offs between them, and on the empirical and the game-theoretic analyses to determine how behavioral and institutional pieces fit together. As we sought to emphasize in developing the details of our account, such determinations should be understood as tools in the service of deliberative democratic theory, so that better institutional prescriptions may be devised to implement more desirable deliberative outcomes.

References

- Austen-Smith, David. 1990. “Information Transmission in Debate.” *American Journal of Political Science* 34(1): 124–52.
- Austen-Smith, David, and Timothy Feddersen. 2005. “Deliberation and Voting Rules.” In *Social Choices and Strategic Decisions: Essays in Honor of Jeffrey Banks*, ed. David Austen-Smith and John Duggan. New York: Springer, 269–316.
- Austen-Smith, David, and Timothy Feddersen. 2006. “Deliberation, Preference Uncertainty and Voting Rules.” *American Political Science Review* 100: 209–18.
- Austen-Smith, David, and William H. Riker. 1987. “Asymmetric Information and the Coherence of Legislation.” *American Political Science Review* 81(3): 897–918.
- Barabas, Jason. 2004. “How Deliberation Affects Policy Opinions.” *American Political Science Review* 98(4): 687–99.
- Battaglini, Marco. 2000. “Multiple Referrals and Multidimensional Cheap Talk.” *Econometrica* 70(4): 1379–1401.
- Bohman, James, and William Rehg. 1996. “Discourse and Democracy: The Formal and Informal Bases of Democratic Legitimacy.” *The Journal of Political Philosophy* 4(1): 79–99.
- Bohman, James, and William Rehg, eds. 1997. *Deliberative Democracy: Essays on Reason and Politics*. Cambridge, MA: MIT Press.
- Calvert, Randall L. 2006. “Deliberation as Coordination Through Cheap-Talk.” Typescript. Washington University.
- Cohen, Joshua. 1996. “Deliberation and Democratic Legitimacy.” In *Deliberative Democracy: Essays on Reason and Politics*, ed. J. Bohman and W. Rehg. Cambridge, MA: MIT Press, 67–92.
- Cohen, Joshua. 1998. “Democracy and Liberty.” In *Deliberative Democracy*, ed. Jon Elster. New York: Cambridge University Press, 185–231.
- Coughlan, Peter J. 2000. “In Defense of Unanimous Jury Verdicts: Mistrials, Communication, and Strategic Voting.” *American Political Science Review* 94: 375–93.
- Crawford, Vincent, and Joel Sobel. 1982. “Strategic Information Transmission.” *Econometrica* 50: 1431–52.
- Dickson, Eric, Catherine Hafer, and Dimitri Landa. 2008a. “Cognition and Strategy: A Deliberation Experiment.” *Journal of Politics* 70(4): 974–89.
- Dickson, Eric, Catherine Hafer, and Dimitri Landa. 2008b. “Rules of Debate: Theory and Experiment.” Typescript. New York University.
- Druckman, James N. 2004. “Political Preference Formation: Competition, Deliberation, and the (Ir)relevance of Framing Effects.” *American Political Science Review* 98: 671–86.
- Dryzek, John S. 2000. *Deliberative Democracy and Beyond: Liberals, Critics, Contestations*. Oxford: Oxford University Press.
- Dryzek, John S., and Christian List. 2003. “Social Choice Theory and Deliberative Democracy: A Reconciliation.” *British Journal of Political Science* 33(1): 1–28.
- Elster, Jon. 1995. “Strategic Uses of Argument.” In *Barriers to Conflict Resolution*, ed. Kenneth Arrow. New York: W. W. Norton, 236–57.
- Epstein, Lee, and Jack Knight. 1998. *The Choices Justices Make*. Washington, DC: CQ Press.
- Estlund, David M. 1997. “Beyond Fairness and Deliberation: The Epistemic Dimension of Democratic Authority.” In *Deliberative Democracy*, ed. J. Bohman and W. Rehg. Cambridge, MA: MIT Press, 173–204.
- Fearon, James. 1998. “Deliberation as Discussion.” In *Deliberative Democracy*, ed. Jon Elster. Cambridge, MA: Cambridge University Press, 44–68.
- Fishkin, James S. 1991. *Democracy and Deliberation: New Directions for Democratic Reform*. New Haven, CT: Yale University Press.
- Gerardi, Dino, and Leeat Yariv. 2006. “Deliberative Voting.” *Journal of Economic Theory* 134: 317–38.
- Glazer, Jacob, and Ariel Rubinstein. 2001. “Debates and Decisions: On a Rationale for Argumentation Rules.” *Games and Economic Behavior* 36: 158–73.
- Gutmann, Amy, and Dennis Thompson. 1996. *Democracy and Disagreement*. Cambridge, MA: Belknap, Harvard Press.
- Habermas, Jürgen. 1984. *The Theory of Communicative Action*, vol. 1. Boston: Beacon Press.
- Habermas, Jürgen. 1990. “Discourse Ethics: Notes on the Program of Philosophical Justification.” In Habermas, *Moral Consciousness and Communicative Action*. Trans. Christian Lenhardt and Shierry Weber Nicholsen. Cambridge, MA: MIT Press, 43–115.
- Habermas, Jürgen. 1996. *Between Facts and Norms*. Cambridge, MA: MIT Press.

- Hafer, Catherine, and Dimitri Landa. 2005. "Deliberation and Social Polarization." Typescript. New York University.
- Hafer, Catherine, and Dimitri Landa. 2007. "Deliberation as Self-Discovery." *Journal of Theoretical Politics* 19(3): 329–60.
- Hardin, Russell. 1988. *Morality Within the Limits of Reason*. Chicago: University of Chicago Press.
- Heath, Joseph. 2001. *Communicative Action and Rational Choice*. Cambridge, MA: MIT Press.
- Jackman, Simon, and Paul M. Sniderman. 2006. "The Limits of Deliberative Discussion: A Model of Everyday Political Arguments." *Journal of Politics* 68(2): 272–83.
- James, Michael Rabinder. 2003. "Communicative Action, Strategic Action, and Inter-Group Dialogue." *European Journal of Political Theory* 2(2): 157–82.
- Johnson, James. 1993. "Is Talk Really Cheap? Prompting Conversation Between Critical Theory and Rational Choice." *American Political Science Review* 87: 74–86.
- Johnson, James. 1998. "Arguing for Deliberation: Some Skeptical Considerations." In *Deliberative Democracy*, ed. Jon Elster. New York: Cambridge University Press, 161–84.
- Knight, Jack, and James Johnson. 1994. "Aggregation and Deliberation: On the Possibility of Democratic Legitimacy." *Political Theory* 22(2): 277–96.
- Knight, Jack, and James Johnson. 2007. "On the Priority of Democracy: A Pragmatist Approach to Political-Economic Institutions and the Burden of Justification." *American Political Science Review* 101(1): 47–61.
- Krishna, Vijay, and John Morgan. 2001. "Asymmetric Information and Legislative Rules: Some Amendments." *American Political Science Review* 95: 435–52.
- Landa, Dimitri. 2005. "The Moral Economy of Deliberative Participation." Typescript. New York University.
- Landa, Dimitri. 2006. "Rational Choices as Social Norms." *Journal of Theoretical Politics* 18(4): 434–53.
- Lax, Jeffrey, and Kelly Rader. 2007. "Procedure and Power: Strategic Opinion Assignment in the Supreme Court." Working Typescript. Columbia University.
- Levy, Gilat. 2007. "Decision-Making in Committees: Transparency, Reputation and Voting Rules." *American Economic Review* 97(1): 150–68.
- Lipman, Bart, and Duane Seppi. 1995. "Robust Inference in Communication Games with Partial Proveability." *Journal of Economic Theory* 66: 370–405.
- List, Christian. n.d. "Group Communication and the Transformation of Judgments: An Impossibility Result." *Journal of Political Philosophy*. Forthcoming.
- Lupia, Arthur. 2002. "Deliberation Disconnected: What It Takes to Improve Civic Competence." *Law and Contemporary Problems* 65: 133–50.
- Macedo, Stephen, ed. 1999. *Deliberative Politics: Essays on Democracy and Disagreement*. Oxford: Oxford University Press.
- Mackie, Gerald. 1998. "All Men Are Liars: Is Democracy Meaningless?" In *Deliberative Democracy*, ed. Jon Elster. New York: Cambridge University Press, 69–96.
- Maltzman, Forrest, James Spriggs, and Paul Wahlbeck. 2000. *Crafting Law on the Supreme Court: The Collegial Game*. Cambridge: Cambridge University Press.
- Manin, Bernard. 1987. "On Legitimacy and Political Deliberation." *Political Theory* 15(3): 338–68.
- Meade, Ellen, and David Stasavage. 2008. "Publicity of Debate and the Incentive to Dissent: Evidence from the U.S. Federal Reserve." *Economic Journal* 118(April): 695–717.
- Meirowitz, Adam. 2006. "Designing Institutions to Aggregate Preferences and Information." *Quarterly Journal of Political Science* 1(4): 373–92.
- Meirowitz, Adam. 2007. "In Defense of Exclusionary Deliberation: Communication and Voting with Private Beliefs and Values." *Journal of Theoretical Politics* 19(3): 301–27.
- Mendelberg, Tali. 2002. "The Deliberative Citizen: Theory and Evidence." *Political Decision Making, Deliberation and Participation* 6: 151–93.
- Miller, N., and Davidson-Podgorny, G. 1987. "Theoretical Models of Intergroup Relations and the Use of Cooperative Teams as an Intervention for Desegregated Settings." *Review of Personality and Social Psychology* 9: 41–67.
- Minozzi, William. 2006. "Ideas for Sale." Mimeo. Stanford Business School.
- Morris, Stephen. 2001. "Political Correctness." *Journal of Political Economy* 109(2): 231–65.
- Mutz, Diana C. 2006. *Hearing the Other Side: Deliberative versus Participatory Democracy*. Cambridge: Cambridge University Press.
- Mutz, Diana C. 2008. "Is Deliberative Democracy a Falsifiable Theory?" *Annual Review of Political Science* 11: 521–38.
- Ottaviani, Marco, and Francesco Squintani. 2006. "Native Audience and Communication Bias." *International Journal of Game Theory* 35: 129–50.
- Patty, John. 2008. "Argument-Based Collective Choice." *Journal of Theoretical Politics* 20(4): 379–414.
- Posner, Richard. 2004. "Smooth Sailing." *Legal Affairs*, January–February.
- Scanlon, T. M. 1998. *What We Owe to Each Other*. Cambridge, MA: Harvard University Press.
- Stasavage, David. 2007. "Polarization and Publicity: Rethinking the Benefits of Deliberative Democracy." *Journal of Politics* 69(1): 59–72.
- Sunstein, Cass, D. Schkade, and Lisa M. Ellman. 2003. "Ideological Voting on Federal Courts of Appeals: A Preliminary Investigation." Public Law Working Paper No 50. University of Chicago.
- Taber, Charles S., and Milton Lodge. 2006. "Motivated Skepticism in the Evaluation of Political Beliefs." *American Journal of Political Science* 50(3): 755–69.