Learning from Debate: Institutions and Information*

Eric Dickson[†], Catherine Hafer[‡], and Dimitri Landa[§]

August 19, 2014

Abstract

We present a model and a laboratory experiment on the informativeness of debate, varying both informational and institutional variables. The informational variable we focus on is a novel factor affecting the extent to which audience members can learn from exposure to unpersuasive arguments. The more easily a listener can learn from an argument she finds unpersuasive, the greater the risk that the speaker will alienate this listener when she fails to persuade her. We find a strong interaction between speakers' responsiveness to that risk and the institutions of debate. When listeners can learn from unpersuasive arguments, many speakers are discouraged from attempting persuasion, irrespective of the debate rules we consider. In contrast, when listeners cannot learn from unpersuasive arguments, debate rules affect speakers' willingness to engage in persuasion.

Keywords: Experiment, rules of debate, learning, argumentation
JEL codes: C9, D8

I. Introduction

In order to make informed and meaningful political decisions, citizens must learn about the issues of the day; their best means of doing so is, often, observing debates between political and media elites. Career policy makers, typically considerably more informed than voters, frequently

*We benefitted from comments from Sandy Gordon, Rebecca Morton, seminar participants at the UC Berkeley, Paris School of Economics, University of Virginia, and Vanderbilt University, and audiences and discussants at the meetings of APSA, MPSA, The Public Choice Society, and SAET.

[†]Corresponding author. Associate Professor of Politics, NYU. Address: 19 West 4th Street, New York University, New York, NY 10012, (212) 992-8697, eric.dickson@nyu.edu

[‡]Associate Professor of Politics, NYU. Address: 19 West 4th Street, New York University, New York, NY 10012, (212) 992-9679, catherine.hafer@nyu.edu

[§]Associate Professor of Politics, NYU. Address: 19 West 4th Street, New York University, New York, NY 10012, (212) 998-3524, dimitri.landa@nyu.edu

rely, in making policy choices, on learning from debates among specialist advisors. At both the mass and the elite levels, the informativeness of such debates is a central determinant of the quality of governance and functioning of democracy. While debates do sometimes deliver in this role, at other times they can be notoriously vacuous, failing to provide policy-makers with the arguments they need to hear, with lamentable consequences, both practical and normative. In this paper, we contribute to the understanding of the issues affecting these outcomes by tracing debaters' incentives in bringing them about. The theoretical arguments and the laboratory experiment we present show that the institutions and the informational setting of debates systematically interact in determining these incentives and the ultimate quality of decision-making.

Consider the following example. Suppose that the only reasonable justification for a new bailout of private banks is that these institutions are "too big to fail." Now suppose that a listener is exposed to an argument that particular banks are indeed "too big to fail," but is not persuaded by it. For some listeners, this outcome may yield little insight beyond the obvious point that that argument is unpersuasive to them: they may lack the necessary understanding of the situation to realize that they should now oppose the bailout because no other justifications for it are reasonable. Other listeners, however, may make this mental leap, and decide to oppose the bailout program – contrary to the speaker's original intent.

This example makes clear that while the communication of an effective argument may further the speaker's cause, it is also possible that arguments for the speaker's cause may *dissuade* listeners, causing them to move away from, rather than toward, the speaker's preferred policy. While the view of political debates as contests in which opponents marshal their best arguments in an attempt to persuade the audience in the direction of their preferred policy choices may be intuitive, effective advocacy when the speakers are uncertain about what listeners may find persuasive, may involve tradeoffs between the hope of persuading them and the danger of alienating them.

The nature of these tradeoffs will depend on the context in which debates take place. Some issue areas are more complex than others; some listeners are more sophisticated than others. The combination of these factors may particularly affect the likelihood with which a listener is alienated by an unsuccessful argument. In settings where a listener is unsophisticated, or where an issue area is particularly complex, the listener is likely to lack the proper context for assimilating new information. As a result, the communication of an unpersuasive argument is likely to have little effect on the listener's position. On the other hand, in settings where a listener is sophisticated, or where an issue area is (for this listener) relatively cut-and-dried, the risk of alienating the listener is likely to be greater.

The bailout argumentation example illustrates an intuition, made more rigorous below, that the risks of alienating an audience may vary depending on the *informational* context. When listeners are sophisticated, when the details of the broader policy debate are commonly understood, when knowledge of the political setting is greater, and when the complexity of the issue area is lower, it becomes likelier that unpersuasive arguments may repel listeners from the speaker's preferred position. These connections, while important for our understanding of political communication, remain largely unexplored.

Further, the *institutional* context in which debates take place may also matter greatly. Speakers' incentives to communicate arguments – or to refrain from doing so out of a fear of alienating listeners – should naturally depend not only on the incentives induced by listeners' qualities, but also on the strategic incentives posed by the presence of opposing speakers. The nature of these strategic incentives may differ, depending on the specific rules governing the debate.

In this paper, we explore the effects of information and institutions on the content of debates using a game-theoretic model and a laboratory experiment. We bring the importance of the informational context into sharp relief by considering two polar-opposite cases: settings in which inferences from unpersuasive arguments are impossible (a *Low Information environment*), and settings in which they are straightforward (a *High Information environment*). While we are not able to manipulate the intelligence or sophistication of particular subjects in the laboratory, we can nonetheless manipulate the "fit" between an issue's complexity and a listener's sophistication by focusing on one setting where *nobody* could *possibly* learn from an unpersuasive argument, and another where such learning should be very easy for nearly everyone.

We also demonstrate the relevance of specific debate rules by considering two distinct alternatives. These alternatives are meant to capture in a stylized fashion a crucial strategic distinction between settings where it is possible to respond to the other speaker and settings where it is not. Under what we refer to as *Simultaneous Speech*, response is not possible. While we model this debate protocol as involving "simultaneous" speeches by two opposed agents, what matters strategically is not that the speeches are literally simultaneous, but rather that a speaker is unable to observe her counterpart's speech before making her own, and is also unable to reply once her counterpart's utterance has become known. Under our second debate rule, *Open-Ended Sequential Speech*, response is always possible: two opposed agents take turns speaking, observing one another's speeches, until both of them consecutively choose to remain silent. Different settings of real-world debate – from closed-chamber deliberations by national political leaders to electoral debates in front of TV audiences to discussions on faculty hires in academic departments – vary in the extent to which speakers can expect to have the right to respond. Our two speech protocols exhibit maximum variation along this dimension.

We find a strong *interactive* effect of the informational condition and the institutional setting. In the *Low Information* environment, the theoretical prediction is that the specific debate rule matters a great deal. As predicted, the risk of alienating the listener holds great sway over speakers under the *Open-Ended Sequential Speech* debate rule, driving out substantive (germane) speech, while under the Simultaneous Speech debate rule, speakers, in contrast, find making germane arguments more attractive and make them considerably more often. This strong effect of debate institutions is conditional, however. Both theory and experimental results suggest that in the *High Information* environment, the risk of alienating the audience drives out germane speech for both *Simultaneous* and *Open-Ended Sequential* Speech, erasing the differences between those rules.

These findings highlight the intertwined importance of both information and institutions in the conduct of debate, but they also speak to one of the key questions in the vast literature on collective decision-making: under what circumstances is it possible for political decision makers (here, the listener) to obtain the information they need to make high-quality decisions? More generally, our findings contribute to a rich tradition of game-theoretic research linking political institutions – the rules under which politics is carried out – to political outcomes. Seminal literatures within political science have explored how voting outcomes can be influenced by details of the voting agenda, as well as how legislative outcomes can be affected by the specific rules under which legislation is considered. Our work extends this approach, and the insight that institutional details matter, to the study of rules of debate.

The remainder of the paper is organized as follows. In Section II, we discuss the related literature on deliberation. Section III introduces the key elements of our formal framework, including definitions of our different rules of debate. In Section IV, we present equilibrium predictions on how the content of debate should be expected to vary across different institutional and informational settings. Section V describes our experimental protocol and presents our experimental results. Section VI concludes. The appendix at the end of the paper comments on the role of some of the assumptions maintained in our analysis.

II. Relation to the Literature

We analyze persuasion in debate as resulting from the communication of arguments that resonate with some segments of the audience. Our theoretical environment relates most closely to the class of persuasion games (Lipman and Seppi 1995; Lanzi and Mathis 2004; Glazer and Rubinstein 2006; Patty 2008) in which the information communicated in senders' messages is (partially) provable. These intrinsically meaningful messages stand in contrast to cheap-talk models of communication (Crawford and Sobel 1982; Austen-Smith and Feddersen 2006; Gerardi and Yariv 2006; Meirowitz 2007), in which the information communicated by the senders' messages is induced by equilibrium beliefs about speakers' strategies. The relative appeal of one rather than the other framework depends on the substantive context (Landa and Meirowitz 2009), but in both types of games, the key question of interest is what determines the senders' incentives to provide information to the receivers.

Within the class of games of persuasion, one of the key distinguishing features of our framework is that the persuasiveness of arguments communicated in debate may be listener-specific: while an argument may be fully persuasive to some receivers, it may be wholly unpersuasive to others. Hafer and Landa (2007) employs this feature but assumes away the possibility of learning from unpersuasive arguments that is a key focus here and studies a different institutional question: the optimal distribution of opportunities to address an audience across speakers with different positions. Glazer and Rubinstein (2006) consider the informational properties of debate rules, including versions of the rules that we analyze below. However, unlike in the present paper, their model assumes arguments are equally persuasive to all listeners, speakers are better informed than listeners about what the latter will find persuasive.

Within the cheap-talk tradition, Austen-Smith (1993) analyzes the effects of different communication rules in a model in which two speakers attempt to influence the actions of a third party. In his model, the speakers are experts whose knowledge of the single-dimensional state of the world – the outcome of a policy – strictly dominates that of the receiver. In contrast, in our environment, the payoff-relevant aspects of the state of the world – the arguments that are convincing to particular receivers – are multi-dimensional, and the senders, though they are in possession of potentially persuasive arguments unavailable to the receivers, are as uninformed about the true state as the receivers themselves. Other work in the cheap-talk tradition also focuses on the effects of the different informational content of speeches. For instance, Minozzi (2011) presents a "jamming theory," in which competing speakers can leave receivers uncertain about who has sent a truthful message. His work analyzes a different pathway of persuasion from that which we consider here; jamming is not possible in our framework, because unlike in Minozzi, speech content is intrinsically persuasive to (some) listeners in our model.

Deliberation and debate are the subject of a considerable experimental and behavioral literature; for a recent survey see Myers and Mendelberg (forthcoming). Our experimental scenario falls within a research tradition that employs an intentionally stylized setting for the study of political communication (e.g., Lupia and McCubbins 1998; Guarnaschelli, McKelvey, and Palfrey 2000; McCubbins and Rodriguez 2006; Dickson, Hafer, and Landa 2008; Minozzi and Woon forthcoming). This kind of setting is a natural fit for exploring the way in which actors' incentives to speak (or not) vary across different institutional and informational conditions because it allows us better to control for agents' prior beliefs as well as for the informational content of speech.

III. The Environment

We analyze four games, each distinguished by its pair of debate rule (*Simultaneous Speech* vs. *Sequential Speech*) and informational environment (*High Information* vs. *Low Information*). We begin our description of these games with the features common to all of them, and then provide precise definitions of the differences between them.

We consider an environment with two speakers, to whom we refer as speaker 0 and speaker 10, and one listener. These speakers make *speeches* during debate; once debate has ended, the listener chooses an *action*.

The speakers' utilities and the listener's utility all depend on which action the listener ultimately chooses. This action is denoted $\pi \in [0, 10]$. The interval $[0, 10]$ can be thought of as a policy space in which each actor has preferences. The two speakers have ideal points at either end of the interval; speaker 0's ideal point, π_0 is 0, while speaker 10's ideal point, π_{10}, is 10. Each speaker knows her own ideal point (and the ideal point of the other speaker). The listener's ideal point, π_L, is either 0 or 10, but neither the listener herself nor the speakers know ex ante which of these values it takes. As such, we model debate as a process through which the listener may learn about her own preferences. The speakers and the listener share a common prior belief about π_L given by $\Pr(\pi_L = \pi_0 = 0) = \Pr(\pi_L = \pi_{10} = 10) = \frac{1}{2}$, assigning equal ex ante probability to the listener's sharing each speaker's ideal point.[1]

The listener's payoff, u_L, and the speakers' payoffs, u_{S_0} and $u_{S_{10}}$, are given by standard

[1]Our theoretical results comparing the consequences of different rules of debate are robust to making priors asymmetric. We adopt a prior equal to 0.5 here both because this yields a presentationally straightforward symmetric game, and because this corresponds to the setup used in our laboratory experiment. We also note that our predictions are similarly robust to changing the value of the constant term in the players' utilities below, as well as to introducing other dimensions on which the speakers may choose to speak.

quadratic utility functions:

$$\begin{aligned} u_L &= 100 - (\pi - \pi_L)^2 \\ u_{S_0} &= 100 - \pi^2 \\ u_{S_{10}} &= 100 - (\pi - 10)^2, \end{aligned}$$

As such, each actor prefers the listener's choice to correspond as closely as possible to that actor's ideal point, and suffers increasing losses the further that choice diverges from that actor's ideal point.[2]

During debate, speaker j chooses her speech(es) s_j from the set $\{\pi_j, x\}$, where, saving on notation, π_j represents an argument that is persuasive for an actor with ideal point π_j, while x represents a non-germane argument (e.g., "this is a great country.") That is, speaker 0 can choose either to make speech "0" – which we will think of as a clinching argument that will be recognized and accepted by agents who share this speaker's ideal point of 0 – or to dodge the question under debate by making an irrelevant utterance.[3] Because a non-germane speech x, unlike a germane speech π_j, remains silent on the topic under debate, we will frequently refer to the choice of x as *silence* and to the choice of π_j as *speaking*.

As noted above, listeners have the potential to learn about their policy preferences over the course of debate. We model listener learning as follows.

If a speaker's speech matches the listener's ideal point (i.e., if speech $s_j = \pi_L$), then the listener learns her ideal point with certainty. The interpretation is that such a matching speech contains an

[2]In the Appendix we comment on the implications of assuming concave utilities in this setting.

[3]We note that our predictions about speaker behavior, as well as our other results, are robust to allowing speakers also to make their counterpart's clinching argument – e.g., for speaker 0 to make the speech "10" – because speakers' interests are never advanced by making such arguments. We restrict the speaker's strategy space here both for clarity in the formal presentation and to match the structure used in our experimental design.

argument that is persuasive to this particular listener, directly triggering recognition of what her policy preferences must be. In this sense, from the perspective of a given listener, persuasiveness is an *intrinsic quality of an argument.* This distinguishes our framework from cheap-talk models of communication, in which particular messages do *not* have intrinsic meanings in this way. In cheap-talk games, listeners can only learn from a given argument indirectly, making inferences based on which speakers are expected to make which arguments in a given equilibrium. In our framework, listeners who hear an argument that is persuasive (to them) can learn directly from the content of a given speech.[4]

If a speaker's germane speech does *not* match the listener's ideal point (i.e., if $s_j = \pi_j$, but $s_j \neq \pi_L$,), we assume that the listener does not automatically learn her ideal point in the same way she learns it when there is a match. The intuition is that an argument that is unpersuasive to a given listener cannot be expected to trigger her optimal policy learning in quite the same way as an argument she finds persuasive. Nonetheless, as previously motivated, an unpersuasive germane speech need not necessarily be uninformative.

In the games we are analyzing, non-germane speech $s_j = x$ is never informative, and the point of the analysis is to understand when speakers prefer to make germane speeches, even at the cost of potentially alienating the listeners, rather than making always uninformative non-germane speeches. To that end, we compare speakers' behavior under two distinct informational conditions that affect the informational content of unpersuasive germane speech, corresponding to the polar opposite cases described in the introduction. Under the assumption of a *High Information* environment, listeners learn just as well from unpersuasive as from persuasive speeches, though

[4]Strictly speaking, given our conception of debate, a listener's ideal point would be thought of as a function of the argument/reason she would find persuasive. To keep exposition simple, we model the effects of speeches on policy preferences directly, rather than employing additional notation to formally represent the set of arguments as distinct from speeches, etc. Bear in mind that potentially informative speeches have this quality because of the specific arguments/reasons they contain.

through different mechanisms. While an unpersuasive speech may not trigger *direct* learning of her preferred policy in the same way that a persuasive speech would, listeners may be able to make indirect inferences about their preferences. If communication takes place in a context that is sufficiently understandable to the listener, such indirect learning can proceed according to the following logic: (1) I just heard speech 0, (2) Speech 0 was not persuasive to me; (3) Being persuaded by speech 0 is necessary for preferring policy 0; (3) Because speech 0 was not persuasive to me, my ideal point must not be 0; (4) If my ideal point is not 0, it must be 10. Of course, this chain of logic requires several distinct kinds of understanding from the listener, including a recognition of the *kind* of speech just heard ("that was speech 0"), of the mapping between arguments and ideal points ("speech 0 is persuasive to all listeners who prefer policy 0"), and of the set of possible ideal points ("if I'm not for it, I must be against it"). In the *High Information* environment, the necessary pieces are in place for listeners to be able to make indirect inferences. As we indicated in the introduction, real-life argumentation settings vary with respect to a number of factors that can be seen to affect the ease of such inferences by the listeners: their own sophistication in understanding the details of the broader policy debate, transparency of the political setting, including the true interests of the debaters, complexity of the issue area, and possibly others. Distinct realizations of these factors induce what is, undoubtedly, a continuum with respect to the ease/difficulty of making inferences from unpersuasive arhuments. Our model of the *High Information* environment is a simple of way of capturing one end of that continuum.

We also analyze debate in the polar opposite case, under the assumption of a *Low Information* environment. In such a setting, listeners cannot learn from an unpersuasive speech either directly or indirectly. This assumption corresponds to a case in which at least one of the links in the above chain of logic is missing; the listener is unable to make sense of her failure to be persuaded by a speech, either because of a lack of proper context or because the issue at hand is too complex relative to her ability to negotiate it.

From a formal perspective, a failure of the first link in the chain – an inability to realize "I just heard speech 0" – can be thought of in at least two conceptually distinct but, in the context of the model, equivalent ways: either the "label" of the speech is unobservable (e.g., an argument was communicated, but the listener is unaware that it should have been the clinching argument for position 0) or the identity of the speaker was unobservable (e.g., the listener is unaware that the speaker was speaker 0, who could only have communicated the clinching argument for position 0). In our experiment (and correspondingly, in our theoretical development), we instantiate this lack of knowledge in terms of a failure to observe the identity of the speaker. The resulting difficulty in learning from unpersuasive arguments is in effect the difficulty in understanding what the unpersuasive arguments *effectively* amount to; the *Low Information* environment models precisely the case in which the meaning of speech is so constrained.[5]

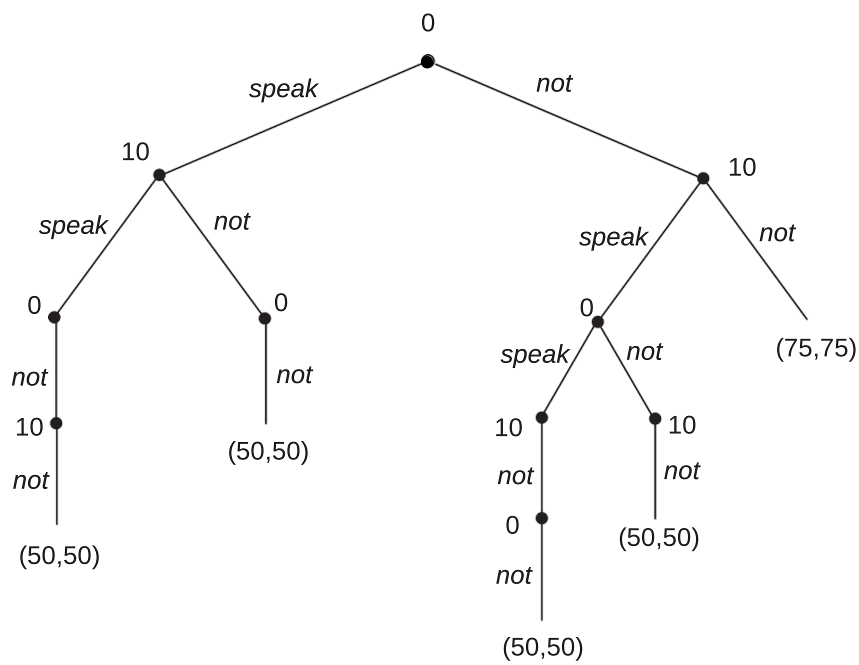
Finally, in each of the two informational environments, we consider two distinct debate rules. Under *Simultaneous Speech*, speakers make one-time simultaneous speeches. It bears repeating that, from the perspective of incentives, what matters is not that the speakers literally speak at the same instant, but rather that each speaker cannot observe the other's choice before making her own. In contrast, under *Open-Ended Sequential Speech*, speakers take turns speaking, each speaker observes all previous speeches and whether the listener has found them persuasive, and the alternation of speakers is repeated until two consecutive choices of "silence," one for each speaker, are made. This game, thus, models the critical feature of sequential open-ended debate: each speaker *always* has an opportunity to respond to a speech by the other (if she has not previously given her own speech herself), so that the identity of the last speaker with the opportunity to make an argument is endogenous to the speakers' debate choices and cannot be known *ex ante*.[6]

[5]Interpreting speaker's identity as a mix of her true preferences and intent, failure to observe the speaker's identity is a natural model of one of the key elements of the listener's political (un)sophistication. A somewhat different interpretation of the *Low Information* environment is that it evokes one of J. S. Mill's classic claims about understanding in a very uncertain world: from the ex ante perspective, it is a model in which, to be sure they understand their ideological position, listeners need to hear different points of view.

[6]The assumption that speakers observe whether the previously made arguments were persuasive to the listener is not consequential for equilibrium play; the differences between the two debate

Note that although open-ended sequential debate can continue to infinity, with each speaker repeating her germane argument over and over again, this is not a repeated game: the payoffs from a stage game here depend on the history, and the game can be terminated as a function of the strategy choices made by each player. Because repeating a germane argument has no effect other than prolonging the game, which, in turn, has no effect on any players' payoffs, we can, without loss of generality, represent Sequential Open-Ended Speech games with the strategically equivalent extensive forms shown in Figure 1 for the *High Information* and in Figure 2 for the *Low Information* environments. In these extensive forms, a given speaker could make a specific speech (e.g., "10") only once; at subsequent information sets involving an opportunity to communicate, this same speaker would have only one alternative, the non-germane message x. Debate then terminates at the first point in time following consecutive choices of x by the two speakers.

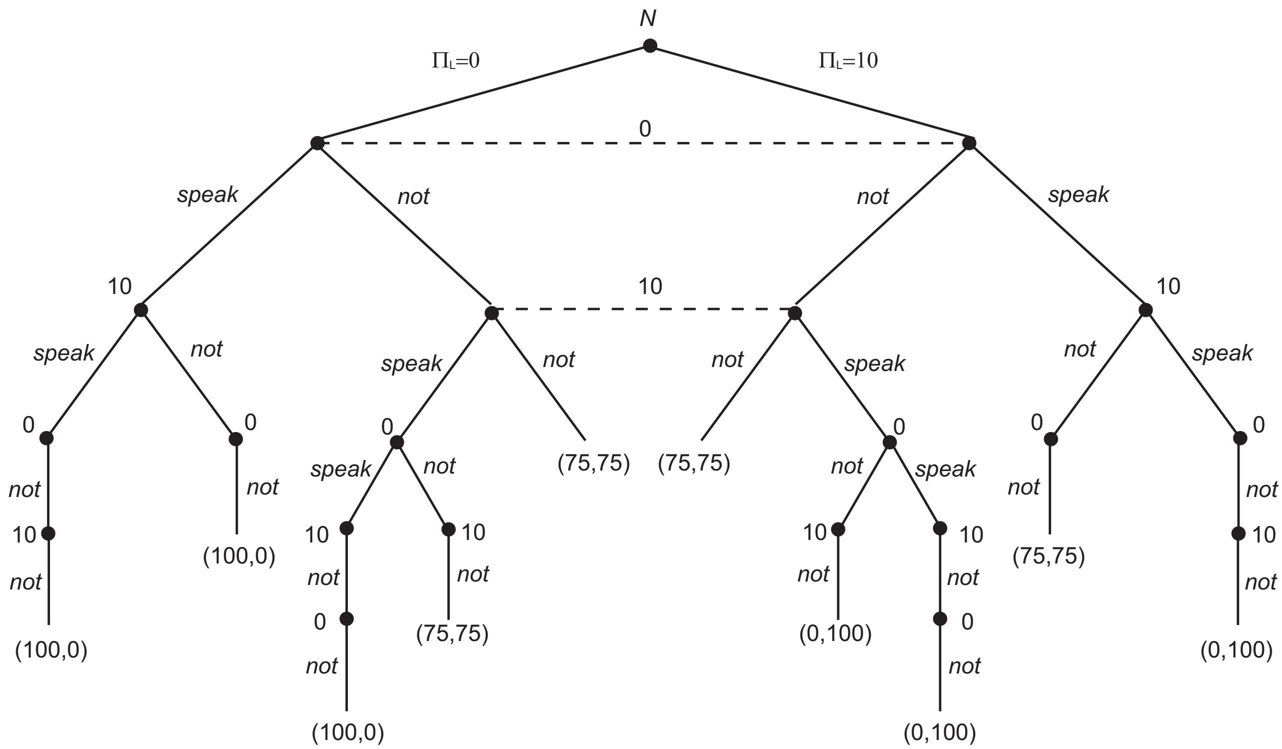
Figure 1: The *Open-Ended Sequential Speech* game in the *High Information* environment.



---

rules come down to the possibility of speakers' responses to each other's actions. The maintained assumption is convenient here because implementing it in the game instantiated in the lab allows for better controlled inferences about subjects' choices in the game.

Figure 2: The *Open-Ended Sequential Speech* game in the *Low Information* environment.



To summarize: crossing our two informational conditions and two debate rules yields four distinct games: (1) *High Information-Simultaneous Speech*, (2) *Low Information-Simultaneous Speech*, (3) *High Information-Open Ended Sequential Speech*, (4) *Low Information-Open Ended Sequential Speech*. We proceed by deriving equilibrium predictions for each of these games, and then continue by describing a laboratory experiment testing these predictions.

IV. Equilibrium Predictions

Below we describe symmetric Perfect Bayesian equilibria. This solution concept requires that strategies be sequentially rational, given beliefs at the time of action. Further, agents are assumed to update their beliefs in response to new information consistent with Bayes Rule. The restriction to symmetric equilibria implies that identical agents should play identical strategies in analytically equivalent situations, regardless of the agents' proper names.

Because the games we consider are symmetric, both speakers face the same incentives for deviation, and thus it is reasonable to suppose that when the speakers' identity is not observable, the off-path-of-play beliefs should assign equal likelihood to each speaker having deviated in the event of deviation. We maintain this specification of the off-path beliefs in what follows and proceed by characterizing the equilibrium predictions first for simultaneous-speech and then for sequential-speech debates.

Simultaneous-Speech Debate

Consider first the case of the *Low Information* environment, in which listeners cannot observe the identity of the speaker. Suppose each speaker's strategy is to speak, and that off-path beliefs are as described above. To see that these strategies and beliefs constitute an equilibrium, consider the possibility of deviation by speaker 0.

Suppose $s_{10} = 10$. If $s_0 = x$,then the listener learns her type with certainty if $\pi_L = 10$, and hence chooses $\pi = 10$, or learns nothing (i.e., her posterior is identical to her prior, $\Pr(\pi_L = 10) = \Pr(\pi_L = 0) = \frac{1}{2}$) if $\pi_L = 0$, and hence chooses $\pi = 5$. Thus, speaker 0's expected payoff from $s_0 = x$, $s_{10} = 10$ is $\frac{1}{2}(100 - (10)^2) + \frac{1}{2}(100 - (5)^2) = 37.5$. Alternatively, if $s_0 = 0$, then the listener learns her type with certainty whether $\pi_L = 10$ or $\pi_L = 0$, and consequently chooses $\pi_L = 10$ or $\pi_L = 0$, respectively. Speaker 0's expected payoff from $s_0 = 0$, $s_{10} = 10$ is $\frac{1}{2}(100-(10)^2)+\frac{1}{2}(100-(0)^2) = 50$. Thus, 0's unique best response to $s_{10} = 10$ is $s_0 = 0$. By symmetry, 10's unique best response to $s_0 = 0$ is $s_{10} = 10$. Thus, both speakers' speaking is an equilibrium. Now consider instead $s_{10} = x$. If $s_0 = x$, then the listener learns nothing and optimally chooses $\pi = 5$, giving speaker 0 a payoff of $(100 - (5)^2) = 75$. Alternatively, if $s_0 = 0$, then the listener learns her type with certainty if $\pi_L = 0$, and hence chooses $\pi = 0$, and learns nothing if $\pi_L \neq 0$, and hence chooses $\pi = 5$. Speaker 0's expected payoff from $s_{10} = x$, $s_0 = 0$ is, then, $\frac{1}{2}(100 - (0)^2) + \frac{1}{2}(100 - (5)^2) = 50 + 37.5$. Thus, $s_0 = 0$ is the unique best response to $s_{10} = x$. By symmetry, 10's best response to $s_0 = x$ is $s_{10} = 10$. Thus, we have the following result[7]:

Result 1 *In Simultaneous Speech debate in the Low Information environment, the unique equilib-*

[7]Note that if one writes speakers' expected utilities in this game as their certain payoffs, then the resulting game would be equilivalent to the classic Prisoners' Dilemma.

rium is both Speakers choosing to speak.

Consider next the case of Simultaneous Speech debate in the *High Information* environment, in which listeners do observe the identity of the speaker. In this case, following speech, the listener knows her ideal point with certainty whether the listener found the speech persuasive or not. Suppose that speaker 10 chooses *speaking*, $s_{10} = 10$. Then the listener learns her ideal point with certainty regardless of her type and speaker 0's action: in terms of our earlier language, the listener learns directly that $\pi_L = 10$ if $\pi_L = s_{10} = 10$ and indirectly that $\pi_L = 0$ if $\pi_L \neq s_{10} = 10$. Given these possibilities, speaker 0 must be indifferent between *speaking* and *silence.* Suppose, though, that speaker 10 chooses to remain silent, $s_{10} = x$. If speaker 0 is silent as well, $s_0 = x$, then the listener learns nothing and chooses 5, yielding a payoff for speaker 0 of $(100 - (5)^2) = 75$. If, however, speaker 0 chooses *speaking*, $s_0 = 0$, then the listener learns her ideal point with certainty and chooses $\pi = 0$ if $\pi_L = 0$ and $\pi = 10$ if $\pi_L = 10$. Speaker 0's expected payoff from $s_{10} = x$, $s_0 = 0$ is $\frac{1}{2}(100 - (0)^2) + \frac{1}{2}(100 - (10)^2) = 50$. Thus, silence is the unique best response to silence.

We have, thus, the following result:

Result 2 *In Simultaneous Speech debate in the High Information environment:*

1. *There are two equilibria: both speakers choosing to speak and both choosing silence;*

2. *The equilibrium in which both are silent Pareto dominates the equilibrium in which both speak.*

Note that there are no mixed strategy equilibria because each player strictly prefers silence to speech if there is a positive probability that her opponent will choose silence.[8]

[8]This point hints at the robustness of the properties of the equilibria in this game to the introduction of expressive benefits from speaking. Let ε represent the expressive benefits from speaking. For $\varepsilon \in [0, 25)$, the two equilibrium profiles in this game remain such, but there is now a third equilibrium, in mixed strategies, in which the expected payoff for each speaker is the same as her payoff in the pure strategy equilibrium in which both speakers speak, which is $50 + \varepsilon$. The payoffs for the equilibrium in which both are silent remain 75, so the equilibrium in which both are

Comparison of Results 1 and 2 makes clear that the information structure matters for the informativeness of debate under the *Simultaneous Speech* debate rule. As we show next, this conclusion does not hold under the *Open-Ended Sequential* debate rule.

Open-Ended Sequential Speech Debate

Because in our framework repeating a speech has no effect on the listener's beliefs (and thus no effect on the listener's action), a speaker in a sequential debate is indifferent between silence and repeating a speech already made. The strategically interesting decision facing a speaker, then, is whether to speak or to be silent when she has not previously spoken. She may be required to make this decision in either of two strategically distinct circumstances: after the other speaker has spoken, or not (whether because the other speaker previously chose silence or because she has not yet had the opportunity to speak).

First consider the *Low Information* environment, and suppose that speaker 10 has already spoken. Does speaker 0 prefer speech or silence? If $\pi_L = 10$, the listener has learned her type and hence chooses $\pi = 10$ regardless of speaker 0's choice. If $\pi_L = 0$, then the listener's choice will depend on s_0. If $s_0 = x$, then the listener learns nothing and hence chooses $\pi = 5$, giving speaker 0 a payoff of 50. Alternatively, if $s_0 = 0$, then the listener learns her type with certainty and chooses $\pi = 0$, giving speaker 0 a payoff of 100. Thus speaker 0's best response to speaker 10's earlier speech is $s_0 = 0$ if $\pi_L = 0$, and both $s_0 = x$ and $s_0 = 0$ if $\pi_L = 10$. By symmetry, speaker 10 also strictly prefers $s_{10} = 10$ in response to 0's earlier speech if $\pi_L = 10$, and both $s_{10} = 10$ and $s_{10} = x$ are best responses to 0's earlier speech if $\pi_L = 10$.

Suppose instead that speaker 10 has not spoken (either because she has not yet had the opportunity, or because she has had the opportunity but chose silence). Does speaker 0 prefer speech or silence? The argument above establishes that if speaker 0 speaks, speaker 10 will subsequently

silent is the Pareto dominant equilibrium for $\varepsilon \in [0, 25)$. Thus, there is no discontinuity between the properties of equilibria in the model with such expressive benefits and the properties of the equilibria described above.

speak. Thus speaker 0 anticipates that making a speech will result in full revelation and, thus, equal chances of the listener choosing 0 and 10, giving speaker 0 an expected payoff of 50. Suppose instead that speaker 0 chooses silence. If speaker 10 then chooses to speak, speaker 0 will subsequently speak also (as established above), resulting in equal chances of the listener choosing 0 and 10 and an expected payoff for each speaker of 50. If, on the other hand, speaker 10 responds to speaker 0's silence with silence, then the debate ends and the listener chooses 5, giving each speaker a payoff of 75. Thus speaker 10 strictly prefers to respond to 0's silence with silence, and so, anticipating that, speaker 0 prefers silence when 10 has not yet spoken. By symmetry, the only equilibrium path of play is one in which both speakers are silent.

Consider now the *High Information* environment. In this case, the listener will learn her type from either possible speech (either directly or indirectly). Her resulting choice is the same regardless of whether she hears speech 0, speech 10, or both speeches 0 and 10. Thus once a speech has been made, both speakers are subsequently indifferent over speech and silence, and furthermore each speaker is indifferent between speaking herself and the other speaker's speaking. Thus the only non-trivial decision facing a speaker is between speech and silence when there has been no previous speech by either speaker. Suppose, then, that speaker 10 has not spoken (again, either because she has not yet had the opportunity, or because she has had the opportunity but chose silence). If $s_0 = 0$, then regardless of the speakers' subsequent behavior, the listener learns her type with certainty, and the expected payoff for each speaker is 50. If $s_0 = x$, then if speaker 10 subsequently speaks, each speaker's expected payoff is 50, and if speaker 10 is subsequently silent, the game ends, the listener chooses $\pi = 5$, and each speaker's payoff is 75. Thus, if $s_0 = x$, speaker 10's best response is to remain silent also, and so speaker 0 prefers $s_0 = x$ to $s_0 = 0$. By symmetry, speaker 10 also prefers $s_{10} = x$ if her opponent has not yet spoken. Thus, the unique equilibrium path of play is one in which both speakers are silent.

We have, thus, the following result:

Result 3 *In Open-Ended Sequential Speech debate, in both High and Low Information environments, there is a unique equilibrium, in which both Speakers choose silence on the path of play.*

The core intuition for this prediction hinges on the importance of counter-arguments. In the debate with Open-Ended Sequential Speeches, one always has a chance to respond to an opponent's speech by making an argument oneself. If a previous speaker moved the audience away from a given speaker's ideal point, then the latter loses nothing, and stands to gain in expectation, by making the opposite argument. But the initial speaker is worse off in expectation if both arguments are made than if no arguments are made – thus deterring her from making her argument. Note that, although no arguments are made on the path of play, the balance of counter-arguments that sustains the no-argument equilibrium here relies on the opposing speaker's credible threat to counterargue off the path of play. The absence of argumentation on the path of play is, thus, a function of the fact that the sequential open-ended speech rule ensures that ability.[9]

V. Experiment

Our experimental design allows us to test two key insights motivated by our theoretical analysis: first, that details of debate rules can greatly influence the amount of information that individuals are willing to communicate during debate; and second, that the extent to which such rules matter can vary as a function of the informational condition.

Experimental Sessions

This section describes data collected from 165 subjects during 10 experimental sessions that were carried out in a social science lab at a large American university. Participants signed up via a web-based recruitment system that draws on a large, pre-existing pool of potential subjects.

[9]Note that under this rule, the introduction of expressive benefits of speaking $\varepsilon \in [0, 25)$ changes off path-of-play behavior in the *High Information* environment so that speaking becomes a unique best response to the opponent's speaking in both information environments, regardless of whether the listener type matched the opponent's speech. However, each player choosing, in turn, to be silent is still the unique equilibrium path of play in the *High Information* case.

Subjects were not recruited from the authors' courses. A filter in the recruitment system blocked subjects from participating in more than one of the experimental sessions (and excluded subjects who took part in our previous lab experiments on deliberation). The subject pool consists almost entirely of undergraduates from around the university.

Subjects interacted anonymously via networked computers. The experiments were programmed and conducted with the software z-Tree (Fischbacher 1999). After giving informed consent according to standard human subjects protocols, subjects received written instructions that were subsequently read aloud in order to promote understanding and induce common knowledge of the experimental scenario. These instructions included screenshots depicting the computer interface that was employed.[10] No deception was employed in our experiment, in accordance with the long-standing norms of the lab in which the experiment was carried out. Before beginning the experiment itself, subjects took an on-screen quiz that both measured and promoted understanding of the instructions.

In each experimental session, subjects interacted with one another over 20 rounds. Each round consisted of one play of a deliberation game, whose game-theoretic structure was closely adapted from one of the games described in Section III. In each round, subjects earned a number of tokens that was determined by the outcome of play; at the end of the experiment, the tokens from three of the twenty periods (randomly chosen by the computer) were converted into dollars (100 tokens = US$7). Subjects' overall payoffs were equal to the sum of payoffs from each of the three randomly-selected periods, plus a US$5 show-up fee.

At the beginning of each round, subjects were randomly assigned to a group of three people; within each group, one subject each was assigned to the roles of "Sender 0," "Sender 10," and "Receiver." "Sender 0" ("Sender 10") was commonly known to have ideal point 0 (10) within a

[10] A sample set of instructions to subjects is included in the Referees' Supplemental Appendix and will be posted online at the time of publication. In this section, terminology from the experimental scenario is introduced in quotation marks where it differs from the theoretical exposition; for continuity, however, the analysis is presented using the terms introduced earlier.

discrete action space $\{0, 1, 2, 3, ..., 10\}$. The receiver's ideal point ("true number") was commonly known to be drawn from $\{0, 10\}$, with each equally likely, but receiver and senders alike were not informed of the outcome of this draw.

In the "communication stage" of each round, speakers had opportunities to communicate that were structured according to one of the particular debate rules to be discussed momentarily. To illustrate the experimental framing of these choices, we note that in our treatment corresponding to *Simultaneous Speech*, "Sender 0" could choose either to send a speech "0" or to send an "empty message" (e.g., x) while "Sender 10" could choose either to send a speech "10" or to send an "empty message." If a given speech (say, "10") matched a receiver's ideal point, the receiver was told the value of the message (in this case, "10"). If a given speech did not match a receiver's ideal point, the receiver was told that he had received an "unmatched message." Finally, if a speaker chose "empty message," the receiver would see "empty message" as the result of communication from that receiver.

In the "final choice stage" of each round, the receiver had an opportunity to choose any point in the action space $\{0, 1, 2, 3, ..., 10\}$. As in our theoretical analysis, the payoffs (in tokens) of receiver and speakers alike involved a quadratic loss function (specifically, 100 minus the square of the distance between the receiver's choice and the subject's ideal point).

At the end of each round, subjects received feedback, including their earnings in tokens for the round, the choice made by the receiver, and what the receiver's actual ideal point was in that round. At the beginning of the next round, subjets were randomly re-matched into a new set of groups and the interaction just described took place anew.

Our laboratory sessions explored behavior within the context of four distinct treatments, involving a two-by-two experimental design. Each of the experimental sessions was devoted to one and only one of the four treatments; each subject was therefore exposed to one treatment only. The first dimension of across-treatment variation involved the debate rule: two of our treatments employed a protocol of *Simultaneous Speech*, while the other two employed a protocol of *Open-Ended*

Sequential Speech.[11]

The second dimension of across-treatment variation involved the extent to which the listener could draw an inference from an "unmatched message." In two of our treatments, set in the *High Information* (*HI*) environment, it was possible for the listener to make such inferences about her true number (her type) from an unmatched message. In these treatments, Receivers were always told the type of the speaker from whom each message had come (e.g., "Sender 0"). As such, receipt of an "unmatched message" following a speech from "Sender 0" ("Sender 10") would mean that the listener's type must be 10 (0) with certainty. In the other two treatments, set in the *Low Information* (*LI*) environment, it was not possible for the listener to infer her true number from an unmatched message, either directly or through knowledge of speakers' equilibrium strategies. We induce this informational condition by *not* informing Receivers of the type of speaker from whom each message had come.[12] Note that our theoretical analysis yields unique equilibrium predictions for the parameter values we employ in our experiment.

The theoretical analysis leads to distinct predictions about speaker behavior on the equilibrium

---

[11] A literal instantiation of the *Open-Ended Sequential Speech* games is susceptible to being hijacked by a subject who chooses to repeat the same message in perpetuity. To avoid this, our experimental instantiation of the *Open-Ended Sequential Speech* protocol follows the strategically equivalent extensive forms depicted in Figures 1 and 2. See the instructions for more information on the framing of our debate protocols to subjects.

[12] Specifically, at the beginning of each round, the labels "the First Sender" and "the Second Sender" were allocated randomly between "Sender 0" and "Sender 10," with each being equally likely to be "the First Sender." Speakers observed the outcome of this allocation of labels, but the receiver did not. As such, receipt of an "unmatched message" from, e.g., "the First Sender" would not allow the receiver to update her beliefs about her own type in the absence of further information. This terminology of "First" and "Second" Senders was used both in the *Simultaneous* and *Open-Ended Sequential* treatments; in the latter, they corresponded to the actual order of speech, so that the "First Sender" moved first.

path for each of our four treatments. In our treatments in the *HI* environment, speakers are deterred from making speeches in equilibrium under *either* of our debate rules. In contrast, in our treatments in the (*LI*) environment, speakers are deterred from making speeches in equilibrium only under our *Open-Ended Sequential* debate protocol, but not under a *Simultaneous* debate rule. Thus, our two-by-two design allows us to test the key insights about the role of debate rules and the informational condition derived from our theoretical model.[13]

Experimental Results: Listeners

Because speakers' realized utilities depend directly on listener behavior, we begin by briefly describing how listeners in our experiment choose actions in the aftermath of debate. This discussion pools data from all four treatments because listeners share the same incentives throughout.

Our theoretical framework predicts that listeners will ultimately choose their action-space ideal point if they learn it during the course of debate. We find that, when subjects learn their ideal point directly, they do in fact choose their ideal point 81.8% of the time (472/577). This figure is high, but falls short of 100%. Almost all of the remaining actions chosen by listeners fall into one of two categories: "5," the midpoint of the policy space (7.6%, 44/577), and an action falling *between* 5 and the listener's ideal point (8.0%, 46/577).[14] A mere 2.6% (15/577) of listeners' choices fell further away than 5 from their ideal point.[15] This pattern suggests that subjects' deviations from

[13]We carried out three sessions of the *LI-Open Ended Sequential* treatment, involving 48 (=21+15+12) subjects; three sessions of the *LI-Simultaneous* treatment, involving 51 (=21+18+12) subjects; two sessions of the *HI-Open Ended Sequential* treatment, involving 36 (=18+18) subjects; and two sessions of the *HI-Simultaneous* treatment, involving 30 (=18+12) subjects.

[14]Of the 46 actions falling between 5 and the listener's ideal point, 18 were one unit away from the listener's ideal point, 12 were two units away, 10 were three units away, and 6 were four units away.

[15]Of the 15 actions falling further away than 5 from the listener's ideal point, 10 were ten units

our theoretical expectations are not driven primarily by misunderstanding of the experimental scenario.

Subjects' answers to our post-experiment debriefing questionnaire shed some light on why some listeners did not choose their own ideal points after learning them in the course of debate.[16] 17% of subjects – a figure quite similar to the 18.2% rate of deviation from equilibrium described above – were classified as making reference to fairness concerns in their debriefing questionnaires. By choosing 3 instead of 0, for instance, a listener with ideal point 0 could increase the payoffs of one of the speakers by 51 tokens, at a cost of only 9 to herself (and to the other speaker). Thus, subjects who value fairness in the form of relative equality of outcomes might be motivated to choose actions that are somewhat biased towards 5 and away from their own ideal points.[17] Such other-regarding behavior resembles that observed in familiar experimental settings such as the dictator game (Forsythe et al. 1994); it is worth noting that the marginal cost of generosity is relatively lower in our setting, because of the concavity of actors' utility functions. Results were similar in the *HI* environment treatments when listeners received only an informative unpersuasive speech (choosing their ideal point 77.1% of the time, 37/48).[18]

---

away (at the opposite pole of the action space), 2 were nine units away, and one each were six, seven, and eight units away, respectively).

[16]An assistant we hired to classify subjects' responses was given our theoretical model and a description of the experimental protocol, so that she would be familiar with the experimental game and the logic of equilibrium in this game, but did not know the results of the experiment before carrying out her classification. She was instructed to classify subjects' responses based on what they might reveal about subjects' motivations or levels of understanding of the experiment, but was given no further direction about the form the classification should take.

[17]As one subject wrote, "I tried to be as fair as possible when submitting a Final Choice so I chose numbers in the middle of the spectrum so that both sides would remain relatively happy."

[18]Our finding that listeners learn about as well from persuasive as from informative unpersuasive speeches may appear to stand in contrast with evidence from other experimental research on

Our theoretical framework also predicts that listeners will choose the midpoint of the policy space (5) when they do *not* learn their ideal point, given their concave utility functions. Indeed, when subjects receive only "empty messages," they do this 69.3% of the time (262/378). Choices of 0 (8.2%, 31/378) and 10 (8.2%, 31/378) are also relatively common. In the classification exercise, 16% of subjects were classified as "gamblers" because of responses alluding to risk-seeking preferences, hot hands, or references to luck and guessing and taking a chance. It appears that risk attitudes likely played a substantial role in motivating many of the subjects who deviated from choosing 5 after failing to learn their ideal point.

Finally, we note that listener behavior appeared to be largely static over time. As an example, consider listener behavior upon learning one's own ideal point "directly" through speech. We noted above that, overall, listeners in this circumstance choose their ideal point 81.8% of the time (472/577). When splitting experimental sessions in half, the corresponding figures are 80.9% in the first 10 periods and 83.1% in the last 10 periods, a difference of only 2.2 percentage points that is not statistically significant ($z = 0.68$, $p = 0.499$, two-tailed difference of proportions test). Listeners who learn their ideal points directly choose either 5 or a number between 5 and their own ideal point at nearly identical rates over time: 15.5% in the first 10 periods and 15.7% in the last 10 periods (a statistically insignificant difference, $z = 0.06$, $p = 0.953$, two-tailed difference of proportions test). Such listeners instead choose actions more distant than 5 slightly less often over

---

deliberation (e.g., Dickson, Hafer, and Landa 2008; on the importance of information processing and cognitive complexity for deliberative outcomes, see Lupia 2002). This contrast should not be pushed too far. The experimental protocol we employ makes the inference from unpersuasive speeches less cognitively demanding than in these studies. It is an intentional design choice that suits our purposes in the present study. We wish to isolate the effects of debate rules as the informational condition varies, and therefore need the ability exogenously to manipulate the extent to which subjects can make such inferences. Nonetheless, it bears noting that the theoretical results we develop above are robust to allowing agents to be inefficient learners, as in Hafer and Landa (2006).

time (3.6% in the first 10 periods, 1.2% in the last 10 periods, a marginally significant difference, $z = 1.75$, $p = 0.081$, two-tailed difference of proportions test). These results suggest that listener behavior is largely invariant over the course of the experimental sessions, with perhaps a slight increase in equilibrium play over time, as subjects gain experience with the experimental scenario.

Overall, these results suggest that the bulk of listener actions were consistent with our theoretical expectations. Most important, this distribution of behavior induces the same incentives in speakers to speak or to remain silent as was the case in our theoretical analyses.

Experimental Results: Speakers in the *Low Information* Treatments

When unpersuasive speeches are uninformative, our theoretical analysis suggests that each speaker will perceive an incentive to send her speech under a *Simultaneous Speech* debate rule, but that each speaker will instead perceive an incentive to send "empty message" under an *Open-Ended Sequential* rule. Figure 3a offers a first glimpse into our data from the *Low Information (LI)* treatments, plotting the fraction of the time that *first* speakers chose to speak over the course of our experimental sessions. Because the second speaker's best responses are conditional on the first speaker's play in the *Open-Ended Sequential* treatment, the data in the graph reflects only the first choice made by the first-moving speaker, to make a better controlled subject-level comparison to the subjects' choices in the *Simultaneous Speech* debate, in which both speakers are effectively first speakers, and there is no such conditionality.[19]

* FIGURE 3 ABOUT HERE *

The data in the figure indicate a considerable across-treatment difference, in the direction predicted by the theory. Averaging over all periods, speakers in the *HI-Simultaneous* treatment chose to send a speech 67.1% of the time (456/680); in contrast, first-moving speakers in the *LI-Open Ended Sequential* treatment chose to send a speech only 34.7% of the time (111/320), a highly

[19]Table 1 will offer a more detailed glimpse into how play unfolds over the *Open-Ended Sequential* extensive form. Figure 4 will show a debate-level comparison of informativeness of debates across institutional rules and information environments.

significant difference ($z = 9.65$, $p < 0.0001$, one-tailed difference of proportions test). Notably, this across-treatment difference became increasingly pronounced as subjects gained experience over the course of the experimental sessions. *LI-Simultaneous* speakers continued to send speeches at a relatively constant rate over time (first five periods: 68.8%; last five periods: 67.1%, a statistically insignificant difference, $z = 0.36$, $p = 0.722$, two-tailed difference of proportions test), while first-moving *LI-Open Ended Sequential* speakers did so less and less often as time progressed (first five periods: 51.25%; last five periods: 20.0%, a statistically significant difference, $z = 4.13$, $p < 0.0001$, two-tailed difference of proportions test). Consistent with visual intuition from the graphs, simple probit specifications do not indicate a significant time trend in the *LI-Simultaneous* data ($z = 0.30$, $p = 0.766$) but do in the *LI-Open Ended Sequential* data ($z = -4.74$, $p < 0.001$).[20]

* TABLE 1 ABOUT HERE *

The data in Figure 3a describe all of the decisions made by speakers in the *LI-Simultaneous* sessions, but, as noted above, describe only the first speaker decision made during each round in the *LI-Open Ended Sequential* sessions. Naturally, speakers' best responses at subsequent information sets are contingent on choices by previous movers. Table 1 offers a more complete depiction of play

[20]These specifications regress the aggregate rate at which speakers send speeches in each round on the round number. In interpreting the time trend in the *LI-Open Ended Sequential* treatment, it is worth noting that subjects performed overwhelmingly strongly on the instructions quiz, indicating that they had good understanding of the experimental instructions before the beginning of play. However, the fact that the subjects understood the rules before play began does not at all imply that they would have fully understood all of the *strategic implications* of the rules at this same point in time. Indeed, it is reasonable to believe that subjects would come to understand these strategic implications primarily within the context of the experience of actually playing the game, as they make decisions in their different, randomly-reassigned roles and receive feedback from one round to the next. As such, it is natural to expect that the predictions of the game-theoretic model would most likely "take" partway through any given experimental session. This intuition is consistent with the time profile we observe in our results.

over the course of the extensive form in the *LI-Open Ended Sequential* sessions. These data offer further evidence of convergence towards our theoretical expectations. Over the course of the *LI-Open Ended Sequential* sessions, when following a first-mover who had *not* made a speech, second-moving speakers sent a speech 32.1% of the time (67/109); this figure decreased from 48.7% in the first five periods to 28.1% in the last five periods (a statistically significant difference, $z = 2.11$, $p = 0.035$, two-tailed difference of proportions test). This indicates a parallel movement in the direction of equilibrium at different decision nodes where actors face similar strategic incentives. In contrast, when following a first mover who *had* made a speech, second-mover behavior diverged sharply depending upon whether that speech had been persuasive. When the prior speech had *not* been persuasive – leaving the listener open to potential persuasion – second-moving speakers chose to make their speech 90.0% of the time (54/60). When the prior speech instead had been persuasive – making the listener a lost cause from the second-moving speaker's perspective – second-moving speakers made their speech only 11.8% of the time (6/51). The relevant figures were comparable for *initially* silent first-moving speakers responding to speech from a second-moving speaker (following an unpersuasive speech, 95.1% of the time (39/41); following a persuasive speech, 19.2% of the time (5/26)).

* FIGURE 4 ABOUT HERE *

The above results have been phrased in terms of speakers' decisions. It is also useful to describe the outcomes in terms of another dependent variable: the fraction of the time that debate is informative, in the sense that the listener receives the information that is necessary for her to learn her own type and so to make a fully informed action choice. Figure 4a plots the fraction of the time that debate is informative over the course of our *LI* treatment sessions. In the *LI-Simultaneous* treatment, debate was informative in this way 65.0% (221/340) of the time, a figure that remained flat over the course of the sessions (first five periods: 67.1%; last five periods: 65.9%; insignificant time trend in a simple probit: $z = -0.05$, $p = 0.961$). In the *LI-Open Ended Sequential* treatment, debate was informative 53.1% (170/320) of the time, a figure that decreased sharply over the course of the sessions (first five periods: 70.0%; last five periods: 42.5%; significant time trend in a simple probit: $z = -3.92$, $p < 0.001$).[21] As subjects become more experienced with the experimental protocol, debate is informative *less* often in the *LI-Open Ended Sequential* treatment than in the *LI-Simultaneous* treatment. Taking individual decisions as the unit of analysis, the overall and last-five-periods figures are both significantly different across treatments at about the $p = 0.001$ level using a one-tailed difference of proportions test.

All of these measures suggest pronounced across-treatment differences consistent with our theoretical predictions. However, as is not uncommon in laboratory tests of institutional differences, our model gets the comparative statics right while erring somewhat in the point predictions. In the *LI-Open Ended Sequential* treatment, subject behavior converges strongly in the direction of our theoretical predictions over the course of the experiment. In contrast, as was evident in Figure 3a, speaker behavior exhibits no time trend in the *LI-Simultaneous* sessions, and speakers make speeches only about two-thirds of the time, as against our 100% theoretical prediction.

The most plausible interpretation of this deviation from our predictions was foreshadowed in the section on listener behavior. As noted before, in the debriefing questionnaire, a substantial number of subjects report that they refrained from making speeches out of concerns for fairness or ethics.[22] A number of these subjects in the role of speaker appear to have been motivated by the fact that mutual silence would guarantee 75 tokens for all three members of their group – assuming that listeners ultimately choose optimal actions given their information – potentially

[21] These specifications regress the aggregate rate at which speakers send speeches in each round on the round number.

[22] One subject wrote: "I always sent no message. I was hoping that all three members of my group would understand the rules of the game in order to have everyone end up with the same amount of money, which would be to always (if you are sender) send the empty message, and have the receiver always choose 5. Then everyone ends up with 75 every time, and ensures that everyone gets an equal good amount of money at the end." One of the more amusing responses concluded: "I chose my false sense of moral superiority over 25 extra tokens. Also I wouldn't want to screw over another empty messager. Keep the dream alive and all that."

offering a substantial improvement over the 0 tokens that a speaker might receive if the listener learned that her ideal point was opposed to that speaker's. This is the case even though the listener benefits from equilibrium play, because listeners always learn their ideal point when both speakers speak, thereby earning 100 tokens instead of an expected maximum value of 75. From the standpoint of external validity, the presence of subject motivations outside those we specifically model is hardly a surprise. However, this is not a particular concern from the perspective of our main analytic goal – the comparison of behavior across different treatments – because such additional motivations would be expected to be present equally in all of our treatments.

Experimental Results: Speakers in the *High Information* Treatments

We now turn our attention to treatments in which unpersuasive speeches are informative. Our equilibrium analyses suggest that debate under *Simultaneous* and *Open-Ended Sequential Speech* should have similar outcomes in settings where unpersuasive speeches in fact are informative, in contrast to the case in which unpersuasive speeches are uninformative.

Figure 3b depicts data from our *High Information (HI)* treatments in a format analogous to Figure 3a. First, note that the pronounced pattern of divergence that was evident in Figure 3a for *LI* sessions is clearly absent from the *HI* data in Figure 3b. Second, averaging over all periods, speakers in the *HI-Simultaneous* treatment chose to send a speech 33.25% of the time (133/400), a rate that is strikingly similar to that for first-mover speakers in the *HI-Open Ended Sequential* treatment, who chose to send a speech 33.3% of the time (80/240) (the difference is not statistically significant, $z = 0.01$, $p = 0.99$, two-tailed difference of proportions test). A simple probit specification regressing the aggregate rate at which speakers send speeches in each round on the round number indicates a significant decrease in the rate at which speakers send speeches both for the *HI-Simultaneous* treatment ($z = -6.95$, $p < 0.001$) and for the *HI-Open Ended Sequential* treatment ($z = -2.76$, $p = 0.006$). As before, as subjects become comfortable with the strategic implications of the rules, behavior comes to resemble the predictions of our model more closely.

Table 2 depicts our *HI* data in the same format as Table 1 did for our *LI* sessions. Once again, the more detailed data offer further support for our theoretical predictions. Over the course of the *HI-Open Ended Sequential* sessions, when following a first-mover who had *not* made a speech, second-moving speakers sent a speech 33.1% of the time (53/160); this figure decreased from 53.1% in the first five periods to 15.9% in the last five periods (a statistically significant difference, $z = 3.45$, $p = 0.0006$, two-tailed difference of proportions test). This pattern conforms closely not only with the first-mover decisions from this treatment, but also with the comparable figures from the *LI-Open Ended Sequential* treatment. As in the *LI-Open Ended Sequential* data, second-mover choices made in the aftermath of a first-mover's speech diverge strongly depending upon whether that speech was persuasive. When the prior speech had *not* been persuasive, second-moving speakers chose to make their speech 88.9% of the time (32/36); when the prior speech instead had been persuasive, second-moving speakers made their speech only 11.4% of the time (5/44). A similar pattern of divergence was observed for *initially* silent first-moving speakers responding to speech from a second-moving speaker (following an unpersuasive speech, 73.1% of the time (19/26); following a persuasive speech, 18.5% of the time (5/27)).

Given these results, it is unsurprising that debate is informative in the sense described above at comparable rates in the *HI-Open Ended Sequential* and *HI-Simultaneous* treatments, and that these rates decrease sharply over the course of experimental sessions.[23] Figure 4b plots the fraction of the time that debate is informative over the course of our *HI* treatment sessions. In the *HI-Simultaneous* treatment, listeners learned their type 50.5% (101/200) of the time, a figure that decreased sharply over the course of the sessions (first five periods: 72.0%; last five periods: 24.0%, a statistically significant difference, $z = 4.80$, $p < 0.0001$, two-tailed difference of proportions test). In the *LI-Open Ended Sequential* treatment, listeners learned their type 55.4% (133/240) of the time, a figure that also decreased sharply over the course of the sessions (first five periods: 75.0%; last five periods: 38.3%, a statistically significant difference, $z = 4.06$, $p < 0.0001$, two-tailed

[23]For the *HI* treatments, we consider debate as being informative either if the listener receives a speech matching her ideal point *or* if she receives an unpersuasive speech, because unpersuasive speeches are informative in these treatments.

difference of proportions test).

VI. Conclusion

Representative democracies cannot properly function in the absence of quality information about policies and their consequences. If public debate among elites is content-free, citizens may retain ill-formed or mistaken views on the issues of the day, and as a result, may cast votes against their own interests or against the public interest. If private debates among advisors fail to inform their principals, the officials elected by citizens may themselves fail to act in the informed best interest of their constituents.

As such, it is vital to understand the circumstances under which debaters do – and do not – choose to reveal their best arguments during debate. Some institutions for debate may give speakers greater incentives to communicate meaningfully than others do; some issues may be more frankly debated than others depending on how complex they are relative to the sophistication of the audience. Understanding how these factors matter can inform the design of institutions for political debate, and ultimately improve the quality of decision making by both political elites and the mass public.

Our paper first addresses these questions by presenting a simple game-theoretic setting that captures key features of debate. We trace speakers' incentives to communicate meaningful arguments as we vary key features of the institutional and informational environment, determining speakers' equilibrium behavior in the context of two distinct debate rules and two distinct informational conditions. Our novel informational variable, the extent to which audience members can learn from exposure to unconvincing arguments, underscores our focus on the tradeoffs speakers face when choosing between revealing and empty speeches.

We then describe the results of a laboratory experiment that evaluates the equilibrium predictions from our game-theoretic model. Our experiment, carried out in the context of a two-by-two design consistent with our theoretical exploration, finds substantial across-treatment differences in speaker behavior – and ultimately in the informational content of debate – according to a pattern that is consistent with our theoretical predictions.

Our results indicate that speakers' strategic incentives are strongly influenced both by institutional and informational factors, and that the informativeness of debate may vary in striking ways across different deliberative conditions. In settings where listeners can learn from unconvincing arguments, speakers' best interests are served by refraining from advancing meaningful arguments at all. Paradoxically, then, to the extent that an audience is aided by debate being as informative as possible, the ability to make inferences from unpersuasive arguments may actually be welfare-reducing for the audience. In contrast, in settings where listeners cannot learn from unconvincing arguments, speakers' incentives depend critically on the specific debate rule that is in place; some rules of debate lead speakers in the direction of maximally informative debate, while others lead speakers in the direction of minimally informative debate.

These findings suggest that the choice of a debate rule may have important normative as well as positive implications – and that the consequences of this choice may vary, depending on the complexity of the issue and the sophistication of audience members. During the course of a debate, audience members may hear arguments that allow them to understand better their own underlying interests, and ultimately, to make better political decisions. However, debate can play such a role only if speakers actually choose to communicate meaningful arguments. Our theoretical and experimental results advance our understanding of how both institutional and informational variables influence the informativeness of debate.

References

[1] Allison, Graham T. 1971. *Essence of Decision: Explaining the Cuban Missile Crisis.* Cambridge: Harvard University Press.

[2] Austen-Smith, David. 1993. "Interested Experts and Policy Advice: Multiple Referrals Under Open Rule." *Games and Economic Behavior* 5, 3-43.

[3] Austen-Smith, David, and Timothy Feddersen, 2006. "Deliberation, Preference Uncertainty,

and Voting Rules." *American Political Science Review* 100 (2), 209-217.

[4] Battaglini, Marco. 2002. "Multiple Referrals and Multidimensional Cheap-talk." *Econometrica* 70 (4), 1379-1401.

[5] Baron, David P. 2003. "Private Politics." *Journal of Economics & Management Strategy* 12 (1), 31-66.

[6] Boudreau, Cheryl and Arthur Lupia. 2011. "Political Knowledge." In *Cambridge Handbook of Experimental Political Science*, James N. Druckman et al, eds., pp. 171-183. Cambridge: Cambridge University Press.

[7] Calvert, Randall L. 2006. "Deliberation as Coordination Through Cheap-Talk." Washington University Typescript.

[8] Crawford, Vincent and Joel Sobel. 1982. "Strategic Information Transmission." *Econometrica* 50, 1431-51.

[9] Dickson, Eric S., Catherine Hafer, and Dimitri Landa. 2008. "Cognition and Strategy: a Deliberation Experiment." *Journal of Politics* 70 (4), 974-989.

[10] Forsythe, Robert, Joel L. Horowitz, N.E. Savin, and Martin Sefton. 1994. "Fairness in Simple Bargaining Experiments. " *Games and Economic Behavior* 6, 347-369.

[11] George, Alexander, L. 1980. *Presidential Decisionmaking in Foreign Policy: The Effective Use of Information and Advice.* Boulder: Westview Press.

[12] George, Alexander L. 1972. "The Case for Multiple Advocacy in Making Foreign Policy." *The American Political Science Review* 66 (3), 751-785.

[13] Gerardi, Dino and Leeat Yariv. 2006. "Deliberative Voting." *Journal of Economic Theory.*

[14] Glazer, Jacob and Ariel Rubinstein. 2001. "Debates and Decisions: on a Rationale for Argumentation Rules." *Games and Economic Behavior* 36, 158-73.

[15] Glazer, Jacob and Ariel Rubinstein. 2006. "A Study in the Pragmatics of Persuasion: a Game Theoretical Approach." *Theoretical Economics* 1 (4), 395-410.

[16] Guarnaschelli, Serena, Richard C. McKelvey, and Thomas R. Palfrey. 2000. "An Experimental Study of Jury Decision Rules." *American Political Science Review* 94, 407-423.

[17] Hafer, Catherine and Dimitri Landa. 2007. "Deliberation as Self-Discovery and Institutions for Political Speech." *Journal of Theoretical Politics* 18 (3).

[18] Hafer, Catherine and Dimitri Landa. 2008. "Majoritarian Debate." New York University Mimeo.

[19] Janis, Irving L. 1972. *Victims of Groupthink: A Psychological Study of Foreign-Policy Decisions and Fiascoes.* Oxford, England: Houghton Mifflin.

[20] Krishna, Vijay and John Morgan. 2001. "A Model of Expertise." *Quarterly Journal of Economics* 116, 747-775.

[21] Lanzi, Thomas and Jerome Mathis. 2004. "Argumentation in Sender-Receiver Games." Mimeographed.

[22] Lipman, Barton and Duane J. Seppi. 1995. "Robust Inference in Communication Games with Partial Provability." *Journal of Economic Theory* 66(2), 370-405.

[23] Lupia, Arthur. 2002. "Deliberation Disconnected: What It Takes to Improve Civic Competence." *Law and Contemporary Problems* 65 (3), 133-50.

[24] Lupia, Arthur, and Mathew D. McCubbins. 1998. *The Democratic Dilemma: Can Citizens Learn What They Need to Know?* Cambridge: Cambridge University Press.

[25] Meirowitz, Adam. 2007. "In Defense of Exclusionary Deliberation: Communication and Voting with Private Beliefs and Values." *Journal of Theoretical Politics* 18 (3).

[26] Minozzi, William and Jonathan Woon. Forthcoming. "Lying Aversion, Lobbying, and Context in a Strategic Communication Experiment." *Journal of Theoretical Politics.*

[27] Myers, Dan C. and Tali Mendelberg. Forthcoming. "Political Deliberation." *Oxford Handbook of Political Psychology*. Oxford: Oxford University Press.

[28] Ottaviani, Marco and Peter Sørensen. 2001. "Information Aggregation in Debate: Who Should Speak First?" *Journal of Public Economics* 81 (3), 393-421.

[29] Ottaviani, Marco and Francesco Squintani. 2006. "Naive Audience and Communication Bias." *International Journal of Game Theory* 35 (1), 129-150.

[30] Patty, John. 2008. "Arguments-Based Collective Choice." *Journal of Theoretical Politics* 20 (4), 379-414.

Appendix

In this appendix, we present some considerations regarding the roles of two assumptions maintained in our analysis: the set of arguments available to each speaker and the concavity of players' utilities.

We first show here that the equilibria in the games described above are robust to letting the set of messages for each speaker be $\{10, 0, x\}$ - that is, to allowing for the possibility of speakers making the other speaker's argument. To demonstrate the robustness of these equilibria, it is sufficient to show that making the opponent's argument is never a unique best response. Because the game is symmetric, we can, without loss of generality, restrict the argument to showing that argument 0 is never a unique best response for speaker 10.

First, note that when the argument labels are observed (that is, when the unpersuasive arguments are informative), speaker 10 must be indifferent between arguments 0 and 10 : in either case, the listener will be able to infer her true type with certainty. Consider, then, the case of uninformative unpersuasive speech under simultaneous-speech debate. If speaker 0 chooses 0, then speaker 10's best response is 10, since that uniquely guarantees support from the listener who does not already know her type with certainty. Suppose next that speaker 0 chooses x, and let w be the listener's choice if unpersuaded by a single speaker, $w \in [0, 10]$. Then, in expectation, speaker 10 will receive $\frac{1}{2}(100 - (0 - 10)^2) + \frac{1}{2}(100 - (w - 10)^2) = \frac{1}{2}(100 - (w - 10)^2)$ if she chooses speech 0 and $\frac{1}{2}(100 - (10 - 10)^2) + \frac{1}{2}(100 - (w - 10)^2) = 50 + \frac{1}{2}(100 - (w - 10)^2)$ if she chooses speech 10. Clearly, speech 10 will dominate speech 0 for all values of w. Suppose, finally, that speaker 0 chooses speech 10. Let $z \in [0, 10]$ be the listener's choice if she is unpersuaded by both speakers. Then speaker 10 will obtain $\frac{1}{2}(100 - (10 - 10)^2) + \frac{1}{2}(100 - (w - 10)^2) = 50 + \frac{1}{2}(100 - (w - 10)^2)$ if she chooses x; $\frac{1}{2}(100 - (10 - 10)^2) + \frac{1}{2}(100 - (z - 10)^2) = 50 + \frac{1}{2}(100 - (z - 10)^2)$ if she chooses 10; and $\frac{1}{2}(100 - (0 - 10)^2) + \frac{1}{2}(100 - (0 - 10)^2) = 50 + \frac{1}{2}(100 - (0 - 10)^2)$ if she chooses 0. Clearly, speech 10 dominates speech 0 for all values of z. Thus, in the simultaneous-speech debate, the speakers will always prefer to make their own arguments to those of their opponents.

Consider next the case of uninformative unpersuasive speech in a sequential-speech debate.

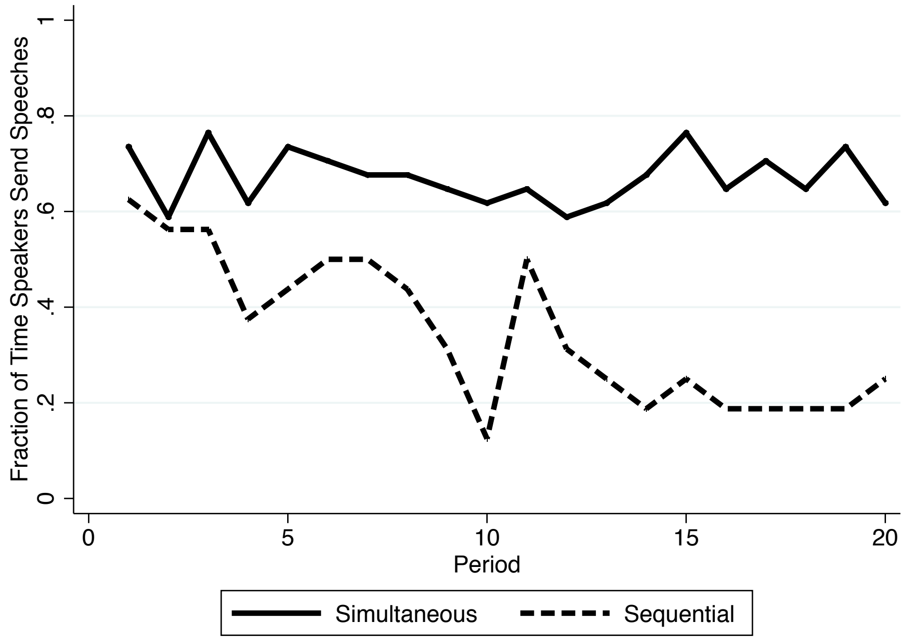
If speaker 0 makes a speech 0, then, similar to simultaneous-speech debate, speaker 10's best response is to make argument 10, since that is the only speech that guarantees that the listener who does not already know their type with certainty will switch to 10. If speaker 0 makes speech x, then choosing anything other that x makes speaker 10 worse off: if she chooses 0, then the listener moves away from her, in expectation, and if she chooses 10, then she provokes speaker 0 to choose 0 next, leading to full revelation, which is *ex ante* inferior to universal silence. If speaker 0 chooses speech 10, then speaker 10's best response is either 10 or x, depending on beliefs after 1 vs. 2 unpersuasive arguments (the argument that speech 10 dominates speech 0 is identical to that above for the simultaneous-speech debate). Thus, speech 0 is never a best response for speaker 10 in sequential-speech debate, and since the games we consider are symmetric, it follows that in all four games we consider, adding the possibility of making the opponent's speech does not upset the robustness of the equilibrium predictions we focus on in the paper.

We next turn to our assumption regarding the shape of individual utilities. Both our theoretical setup and the experiment based on it posit concave (quadratic) payoffs for the players – that is, assume that players enjoy a decreasing marginal utility in proximity of the outcome to their type, or, alternatively, that they are risk-averse. It is easy to see that the equilibrium predictions of our model will vary with this assumption. If the listener's payoffs are convex, then she will always prefer to choose either 1 or 10 to any interior value on the policy choice interval. If so, then one of the speakers (whose preferred policy would not be chosen by the listener) will always have an incentive to speak, and if so, then, assuming any residual uncertainty, so will the other speaker. Thus, assuming that the listener's payoff is convex, the equilibrium outcome across the debate rules and informational environments will always be full revelation. The same outcome will occur if the speakers' payoffs are convex as well, since they will now prefer the expected value of full revelation to that of the listener's choice under complete ignorance.

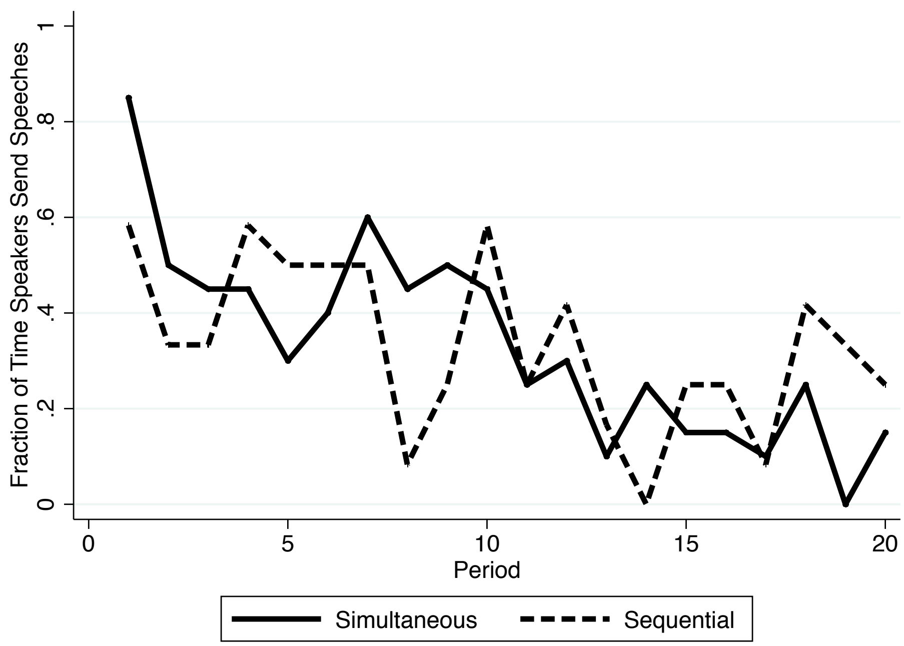
We focus our analysis on the case with concave utilities for several reasons. First, this is the assumption that best fits with the bulk of the political science and economics literature on policy choices in spatial settings, and so facilitates comparability and interpretation of our results in light of the previous studies. Second, to the extent that we think that attitudes toward risk systematically vary across circumstances, our approach in this paper is consistent with the interpretation of the experimental results as establishing behavioral possibilities rather than as causal explanations of behavior observed outside the laboratory. Third, unlike convex payoffs, concave payoffs generate variation in predictions, and to the extent that one would like to be able to use our findings to understand variation in outcomes, variation in predictions is essential. Finally, one of our motivations is normative: to understand how and when we can implement fully-revealing debate. In assuming concave payoffs, we are working with the harder case for realizing that outcome - arguably the right approach from the standpoint of developing a theory of robust institutional design.

Figure 3: Frequencies With Which Speakers Made Speeches (by Round).

(a) *Low Information* Treatments



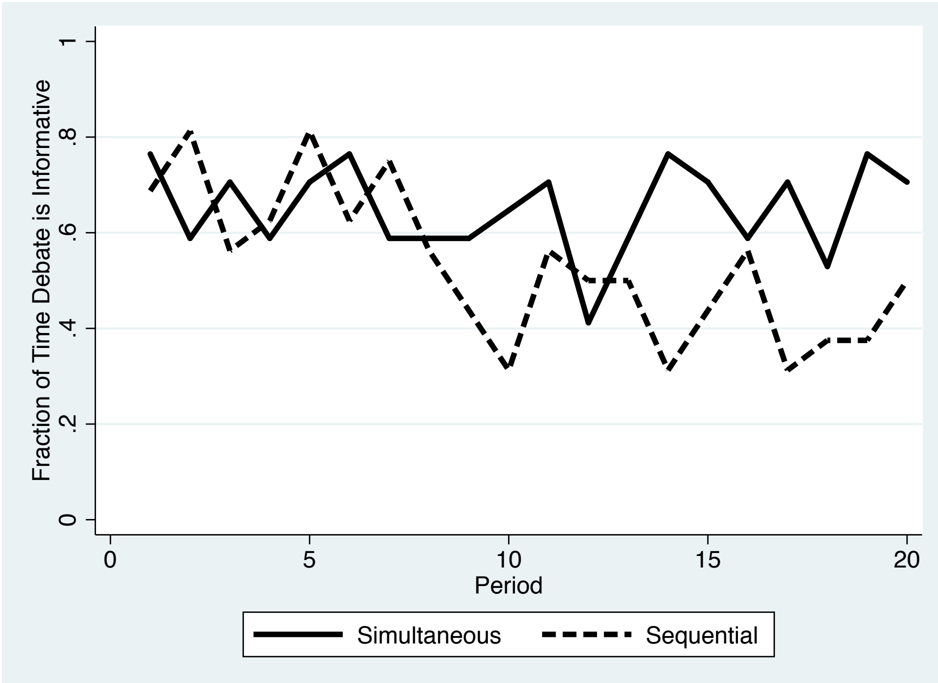
(b) *High Information* Treatments



Note: The Sequential data is from the first speaker's first move only.

Figure 4: Frequencies With Which Debate Is Informative (by Round).

(a) *Low Information* Treatments



(b) *High Information* Treatments

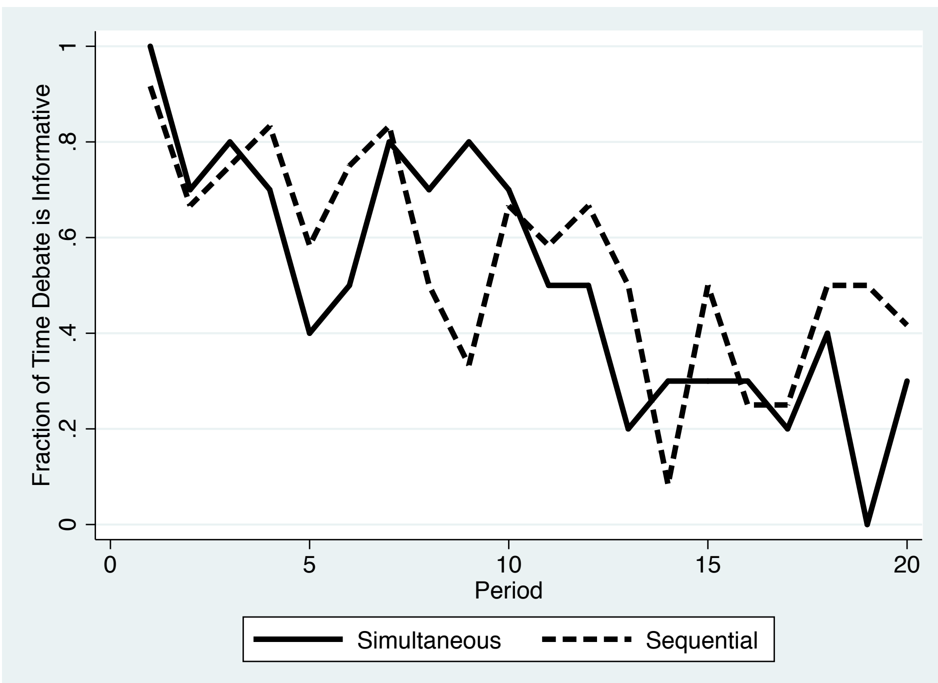


Table 1. Frequencies With Which Speakers Made Speeches: *Low Information* Treatments

	Simultaneous	O.E. Sequential 1st Speaker at beginning	O.E. Sequential 2nd Speaker if 1st Speaker silent	O.E. Sequential 2nd Speaker if 1st Speaker unpersuasive	O.E. Sequential 2nd Speaker if 1st Speaker persuasive	O.E. Sequential 1st Speaker if 2nd Speaker unpersuasive	O.E. Sequential 1st Speaker if 2nd Speaker persuasive
all pds	67.1% (456/680)	34.7% (111/320)	32.1% (67/209)	90.0% (54/60)	11.8% (6/51)	95.1% (39/41)	19.2% (5/26)
pds 1-5	68.9% (117/170)	51.25% (41/80)	48.7% (19/39)	88.9% (24/27)	21.4% (3/14)	88.9% (8/9)	30.0% (3/10)
pds 6-10	66.5% (113/170)	37.5% (30/80)	28.0% (14/50)	94.4% (17/18)	8.3% (1/12)	100.0% (12/12)	0.0% (0/2)
pds 11-15	65.9% (112/170)	30.0% (24/80)	28.6% (16/56)	77.8% (7/9)	6.7% (1/15)	90.9% (10/11)	20.0% (1/5)
pds 16-20	67.1% (114/170)	20.0% (16/80)	28.1% (18/64)	100.0% (6/6)	10.0% (1/10)	100.0% (9/9)	11.1% (1/9)
theoretical prediction	100%	0%	0%	100%	(indiff.)	100%	(indiff.)

Note: O.E. Sequential = Open Ended Sequential

Table 2. Frequencies With Which Speakers Made Speeches: *High Information* Treatments

	Simultaneous	O.E. Sequential 1st Speaker at beginning	O.E. Sequential 2nd Speaker if 1st Speaker silent	O.E. Sequential 2nd Speaker if 1st Speaker unpersuasive	O.E. Sequential 2nd Speaker if 1st Speaker persuasive	O.E. Sequential 1st Speaker if 2nd Speaker unpersuasive	O.E. Sequential 1st Speaker if 2nd Speaker persuasive
all pds	33.25% (133/400)	33.3% (80/240)	33.1% (53/160)	88.9% (32/36)	11.4% (5/44)	73.1% (19/26)	18.5% (5/27)
pds 1-5	51.0% (51/100)	46.7% (28/60)	53.1% (17/32)	72.7% (8/11)	23.5% (4/17)	100.0% (7/7)	30.0% (3/10)
pds 6-10	48.0% (48/100)	38.3% (23/60)	37.8% (14/37)	90.9% (10/11)	8.3% (1/12)	71.4% (5/7)	0.0% (0/7)
pds 11-15	21.0% (21/100)	21.7% (13/60)	31.9% (15/47)	100.0% (6/6)	0.0% (0/7)	55.6% (5/9)	16.7% (1/6)
pds 16-20	13.0% (13/100)	26.7% (16/60)	15.9% (7/44)	100.0% (8/8)	0.0% (0/8)	66.7% (2/3)	25.0% (1/4)
theoretical prediction	0%	0%	0%	100%	(indiff.)	100%	(indiff.)

Note: O.E. Sequential = Open Ended Sequential