Rules of Debate: Theory and Experiment

Eric S. Dickson* Catherine Hafer† Dimitri Landa‡

September 22, 2008

Abstract

We present a game-theoretic model of debate and a laboratory experiment that explore how strategic incentives to make potentially persuasive arguments vary across different informational and institutional contexts. In our model, a key feature of the informational environment is the extent to which members of a debate audience are able to extract informational content from exposure to an argument that they find unconvincing. Our theoretical results show that when the informational content of unconvincing arguments is relatively high, speakers are discouraged from arguing irrespective of the distinct institutional rules of debate that we consider. By contrast, when the informational content of unconvincing arguments is relatively low, debate rules matter: speakers may be lead towards maximally or minimally informative debate, depending on the debate rule. In a laboratory experiment, we vary the informational and institutional settings for debate across four distinct treatments, and observe patterns of behavior which are broadly consistent with the predictions of our model.

*Assistant Professor of Politics, NYU, e-mail: eric.dickson@nyu.edu

†Associate Professor of Politics, NYU, e-mail: catherine.hafer@nyu.edu

‡Assistant Professor of Politics, NYU, e-mail: dimitri.landa@nyu.edu

I. Introduction

Moral and political debates are often thought of as contests in which opponents marshal their best arguments in an attempt to sway the audience in the direction of their preferred alternatives. While this conception of debates is both evocative and appealing, the informational, institutional, and political context of debates can vary dramatically from one setting to another. To what extent do the contents of debate vary across these different environments? Using a game-theoretic model of debate and a laboratory experiment testing insights derived from the model, this paper analyzes the ways in which debaters' deliberative decisions – in particular, their willingness to engage in informative argumentation – are affected by the specific rules under which debate is carried out.

We model persuasion as resulting from the communication of arguments that resonate with some segments of the audience. In our framework, individual members of the audience make choices in the aftermath of debate. Both the "speakers" (debaters) and the audience members themselves have a stake in the choices that are made. Different individuals within the audience, who find different arguments to be convincing, have different optimal choices. Debate can potentially affect the choices of audience members by presenting to them arguments which they may (or may not) find compelling, thereby causing them to update their beliefs about which choices are ultimately better or worse for them personally. Speakers' stakes in audience members' choices induce strategic incentives for speakers to choose their arguments in a way that maximizes the expected proximity of listeners' choices to the choices speakers themselves would prefer to be made on the issue at hand.

The game-theoretic and experimental results we present suggest that different features of the institutional and informational environment in which debate takes place have profound consequences for the deliberative decisions that speakers can be expected to make. Strikingly, the specific rules under which debate is conducted can have a profound impact on the extent to which debate is informative – that is, the extent to which speakers are willing to express their arguments to the audience at all.

We explore how strategic incentives vary across both different rules of debate and different informational environments. A key feature of the informational environment that affects those incentives is the extent to which each given member of the audience is able to extract informational content from exposure to an argument that he or she finds *un*convincing. That is, in some settings, the structure of information allows audience members to learn that they should update *away* from the speaker's preferred position upon receipt of an unconvincing argument, while in other settings, the structure of information permits no such inference whatsoever. It is perhaps unsurprising that such a difference in the informational environment may have some consequences for speakers' strategic incentives. However, what we establish in our game-theoretic analyses, and demonstrate in the laboratory, is much more subtle: namely, that such features of the informational environment have a profound impact on incentives under some rules of debate, but no impact under others. Specifically, we find that when the informational content of unconvincing arguments is relatively high, this discourages argumentation by speakers across different debate rules. In contrast, when the informational content of unconvincing arguments is relatively low, speakers' incentives across debate rules diverge sharply; in particular, 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 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 can matter, to the study of rules of debate.

The remainder of the paper is organized as follows. In Section II, we discuss related literature on deliberation. Section III introduces the key elements of our formal model, including definitions of our different rules of debate. In Section IV, we present equilibrium analysis of our formal model that offers predictions as to the way in which the contents of debate should be expected to vary

across different institutional and informational settings. Section IV also discusses the robustness of our theoretical results to alternative specifications. Section V describes our experimental protocol and presents our experimental results. Section VI then concludes. Proofs of our formal results appear in an appendix.

II. Relation to the Literature

The existing formal literature on deliberation focuses, by and large, on the analysis of cheap-talk models of information transmission. In such models, individuals possess private information, and send messages (which may or may not be accurate) about the contents of this information; within limits, deceptive messages have the potential to be convincing in equilibrium. In cheap-talk models, the credibility of any given message is endogenous to the strategic interaction between the message's sender and its receiver. Whether a message is credible in equilibrium depends on the proximity of the sender's primitive preferences to those of an individual receiver (Crawford and Sobel 1982), or to those of an expected majority within a group of receivers (Meirowitz 2006, Meirowitz 2007), as well as on the sender's being pivotal both as a source of information and as a voter (Austen-Smith and Feddersen 2006).

Lipman and Seppi (1995) and Lanzi and Mathis (2004) analyze a sender-receiver model in which senders can supply partial proofs for their signals and in which the *veridicality* (truth content) of a message is the same for all types of receivers. Common veridicality is also assumed in Glazer and Rubinstein (2006), who analyze a model of persuasion in which the messages contain "hard" or fully provable information, but this information is partial and the informativeness of messages is a function of speaker credibility. In contrast, in the model we analyze, messages are fully provable and complete insofar as they are convincing, but their veridicality differs across agents, and when the proof is rejected (that is, when they are not convincing), they may still have informational content that is either exogenously fixed or, as in the cheap-talk models, endogenously derived from the equilibrium incentives of the players. Common veridicality is also assumed in Patty (forthcoming),

but unlike in the present paper, players' preferences need not be driven by the propositional content of arguments they find convincing.

The closest models to the one developed in the present paper are Glazer and Rubinstein (2001) and Hafer and Landa (2007, 2008). The latter analyze related models with argumentation that, like in the present model, satisfies full provability and private veridicality but, unlike the current model, impose exogenously fixed degenerate distribution on the meaning of unpersuasive arguments. Also, unlike the current model's focus on the properties of debate rules, their focus is on the informational effects of institutions affecting the speaker's access to the audience and of preference aggregation rules, respectively. Glazer and Rubinstein consider the informational properties of debate rules, including simultaneous speech and a version of the fixed-end sequential speech rules that we analyze below. Unlike the model in the present paper, their model is one of common veridicality between the speakers and the (single) listener and assumes that speakers are better informed than the listener about what the latter will find persuasive.

There are number of papers that consider properties of communication with multiple senders within a cheap-talk environment. Within this literature, Austen-Smith (1993) and Krishna and Morgan (2001) analyze the effects of different communication rules in a model in which two speakers attempt to influence the actions by a third party. Battaglini (2002) examines a model with a multidimensional state variable. Ottaviani and Sørensen (2001) analyze the effects of orders of speech on committee debate with speakers who have heterogenous expertise and career concerns. Gerardi and Yariv (2006) explore the relationship between voting rules and communication protocols in the context of a Condorcet Jury Theorem environment with pre-play communication. Calvert (2006) considers the coordination effects of pre-play cheap-talk communication without policy uncertainty.

Deliberation and debate are the subject of a considerable experimental and behavioral literature; Mendelberg (2002) provides a comprehensive review of research up to that date. Our experimental scenario falls within a research tradition that employs a 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). Our research explores the way in which actors' incentives to speak (or not) vary across different institutional and informational conditions; a stylized setting is a natural fit for such a research question, because it allows us better to control for agents' prior beliefs as well as the informational content of speech.

III. The Model

Let the population of agents consist of the set of senders/speakers $\mathcal{S} = \{1, 2\}$, and the set of receivers/listeners $\mathcal{R}$, where $|\mathcal{R}| \geq 1$ and finite. For simplicity, suppose that $\mathcal{S} \cap \mathcal{R} = \emptyset$. The basic sequence of the interactions we analyze includes first public debate between the senders and then choices by the receivers. We adopt the convention of referring to receivers' choices as *actions* and to the senders' choices as *speeches*.

We assume that only receivers make action choices: each $i \in \mathcal{R}$ chooses $\pi_i \in [0, 1]$. However, both senders and receivers have ideal points in the action space. In particular, for $i \in \mathcal{S} \cup \mathcal{R}$, $\hat{\pi}_i \in \{0, 1\}$ is i's ideal action, which we will refer to as i's type. For simplicity, we assume that the speakers have no uncertainty over their own ideal points. In order to operationalize the receivers' uncertainty in a manner consistent with the notion of debate discussed above, suppose that receiver i's ideal action choice is a function of the reason/argument she finds convincing. In particular, for each receiver i, let $r_i \in \{0, 1\}$ be such an argument and let $\hat{\pi}_i = r_i$. A receiver i's uncertainty about r_i then implies her uncertainty about $\hat{\pi}_i$. Let $\theta \in [0, 1]$ be the common prior belief about i's type for each $i \in \mathcal{R}$, where $\theta = \Pr(\hat{\pi}_i = 1)$. The type for each i is an independent draw from this distribution. Debate between the senders may lead to i's learning her unknown value of r_i, and consequently, the corresponding value of $\hat{\pi}_i$.

We assume that each agent's payoffs depend directly on the actions of the members of the society as well as her own (cf. Baron 2003).[1] Because the receiver cannot affect the actions of others and

[1]Examples of issues which would most immediately fit this description include the desirability of obtaining abortions (holding constant the official policy); the work-force participation of women; the value of post-secondary edu-

hence her utility from their actions, for the purpose of determining equilibrium behavior, her utility can be effectively described as being over her own actions alone. In particular, for all $i \in \mathcal{R}$,

$$u_i(\hat{\pi}_i, \pi_i) = -(\pi_i - \hat{\pi}_i)^2.$$

Because speakers only choose speeches, and not actions, each speaker i's direct utility is defined only over the receivers' actions. In particular, for all $i \in \mathcal{S}$,

$$u_i(\hat{\pi}_i, \pi_i) = -\sum_{j \in \mathcal{R}} (\pi_j - \hat{\pi}_i)^2.$$

The substantive implications of our results extend to any concave loss functions for speakers and receivers.

For each speaker i, a speech s_i is a choice from a set $\{\hat{\pi}_i, x\}$, where a choice of $\hat{\pi}_i$ represents a decision to communicate the argument i herself finds convincing (which, in our model, is equal to $\hat{\pi}_i$),[2] and where x can be interpreted as silence or as an uninformative or irrelevant remark, e.g., "this is a great country." (In the discussion of robustness in Section IV, we consider the possibility of defining $s_i \in \{0, 1, x\}$.) In what follows, we reserve the term "argument" only for those statements that are potentially persuasive (that is, 0 or 1, but not x), and use the shorthand "speech" or "speaking" to refer to an instance in which a speaker chooses to make an argument, and "no speech" "not speaking" or "silence" to an instance in which the speaker chooses x.

If receiver i hears a speech that corresponds to her argument, $s_j = r_i$, then she learns r_i, and thus $\hat{\pi}_i$. In this sense, from the perspective of a particular receiver, persuasiveness is an intrinsic quality of an argument; it does not depend on the identity or preferences of the individual who articulated it. If, on the other hand, the receiver hears speech that is not persuasive to her, modeled

---

cation; choices concerning child-bearing, etc. Our formulations of individual preferences could also be thought of as an approximation to the context in which a single binding decision is made following deliberation.

[2]We note that, from a formal perspective, the inclusion of r_i is somewhat redundant, in that the model could be posed in terms of $\hat{\pi}_i$ alone. We specify the model as we do, including r_i, better to communicate the intuition underlying our model of persuasion.

here as $s_j \neq r_i$, then she does not learn r_i directly but may still be in a position to make inferences about the equilibrium value of s_j and thus to update her belief about $\hat{\pi}_i$ accordingly.

In order to capture the notion that receivers may be able to make inferences about their own ideal points even from arguments that they find unpersuasive, we introduce two parameters, corresponding to distinct grounds for such inferences. The first parameter, q, captures a variety of exogenous indicators of informativeness. In particular, we assume that when $s_j \neq r_i$, receiver i may still "observe the value of s_j" with probability q. Thus when the receiver is not directly persuaded, she is nonetheless sometimes able (i.e. with probability q) to assess directly the informational content of s_j and to determine that $r_i = 1 - s_j$ with certainty.

With probability $(1-q)$, however, a receiver i who is not persuaded by the speech s_j (that is, for whom $s_j \neq r_i$) cannot make inferences about r_i by relying on exogenous factors. As in the standard cheap-talk model, she may nonetheless be able to use her knowledge of equilibrium play and her beliefs about her own and the senders' preferences to assess r_i indirectly. In the general case we consider, the ideal points of the senders are private information and all $i \in \mathcal{R}$ assign a common prior probability $\varphi \in [0, 1]$ to any given speaker $j \in \mathcal{S}$ being of type $\hat{\pi}_j = 1$, $\varphi = \Pr(\hat{\pi}_j = 1)$. The type for each j is an independent draw from this distribution. (In the discussion of robustness in Section IV, we consider the possibility of speaker-specific φ.) Given the conjecture of equilibrium play, the value φ provides a second, and distinct, source of information about r_i. In particular, conditional on knowing that senders of type $\hat{\pi}_j$ make speeches $s_j(\hat{\pi}_j)$ in equilibrium, the value of φ determines how much information the receiver can obtain about her own type $\hat{\pi}_i$ from the fact that she found the speech unconvincing.

Debate Rules

A general specification of a *debate protocol* is as a triple consisting of a set of speakers $\mathcal{S}$, a set of speaker messages $\{\hat{\pi}_1, x\} \times \{\hat{\pi}_2, x\}$, and a debate rule. A *debate rule* is defined by a pair consisting of a speaking order and a debate termination protocol.

A speaking order O is an ordered list of speakers, o elements long, such that $o \geq |\mathcal{S}|$ and $\forall i \in \mathcal{S}$,

there is $j \in \{1, ..., o\}$ s.t. the jth element of the list O, O^j, is i. Thus, the definition of the speaking order requires that each member of $\mathcal{S}$ appears on it at least once. The *debate termination protocol* specifies how the debate ends relative to the speaking order.

We consider three different debate rules:

Definition 1 *The debate rule is* ***simultaneous speech*** *if the debate consists of and ends with one-time simultaneous speeches by all members of* $\mathcal{S}$*.*

Of course, the substantive importance here is not that the speakers literally speak at the same time, but that they do not observe each other's speeches until both speakers are done. Simultaneous speech debate rules are common in print media - e.g., left vs. right columns on the editorial pages of newspapers - but also in electoral politics when candidates have to commit to a long-term campaign strategy and substantial revisions to it are infeasible.

Definition 2 *The debate rule is* ***open-ended sequential speech*** *if (1) each speaker i has an opportunity to make a speech s_i in the sequence specified by the speaking order, and each speaker hears all previous speeches before making her own; and (2) the speaking order is repeated until no speaker strictly prefers to continue the debate.*

Formally this debate rule can be described as follows. Given a speaking order O, at time t, speaker O^t speaks and then if any $i \in \mathcal{S}$ prefers to speak, the debate continues; otherwise, the debate is closed. If $t \in \{o+1, ..., 2o\}$, the speaker in t is O^{t-o}; if $t \in \{ko+1, ..., (k+1)o\}$ where $k \in \{1, 2, ...\}$, then the speaker in t is O^{t-ko}.

This debate rule is closely related to the next, fixed-end sequential speech rule, which differs from it in its debate termination protocol. Whereas the open-ended sequential speech rule has an endogenous termination, the fixed-end sequential speech rule has an exogenous termination:

Definition 3 *The debate rule is* ***fixed-end sequential speech*** *if (1) each speaker i has an opportunity to make a speech s_i in the sequence specified by the speaking order, and each speaker hears all*

previous speeches before making her own; and (2) the debate ends immediately after the last speaker in the speaking order makes her speech.

In our formalism, given a speaking order O as defined above, the fixed-end sequential speech rule requires that at time $t = 1$, speaker O^1 speaks, at $t = 2$, O^2 speaks, etc. After O^o speaks, debate ends.

Both types of sequential speech debate rules are particularly common in assemblies. Note, however, that their definitions allow for the possibility that different speakers could appear in the speaking order unequal as well as equal numbers of times (but, as the above definition of a speaking order requires, no less than once). As our results show, how many times a speaker appears in a sequential-speech debate turns out to have no effect on the debate's equilibrium informational content, although under some conditions the order in which they appear does.

Finally, we assume that under the sequential speech rules, speakers observe all previous speeches, as well as the extent to which these speeches were persuasive, in the histories prior to their turns. While the observation of previous speeches themselves is important, observation of audience responses to previous speeches matters only off the path of play and does not affect behavior on the path.

IV. Equilibrium Analysis

Throughout, our equilibrium concept is Perfect Bayesian equilibrium in strategies with undominated actions. It requires that strategies chosen include only undominated actions and 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. We restrict our analysis in what follows to symmetric equilibria - that is, equilibria in which identical agents play identical strategies in analytically equivalent situations.

For all receivers $i \in \mathcal{R}$, i's expected utility is given by

$$E[u_i(\hat{\pi}_i, \pi_i)|\Pr(\hat{\pi}_i = 1)] = -[\Pr(\hat{\pi}_i = 1|\cdot)(\pi_i - 1)^2 + (1 - \Pr(\hat{\pi}_i = 1|\cdot))(\pi_i)^2].$$

For all $i \in \mathcal{R}$, the expected utility maximizing action choice $\pi_i = \Pr(\hat{\pi}_i|\cdot)$. Receivers' beliefs following speech can be characterized as follows. If $s_j = r_i = 1$, then $\Pr(\hat{\pi}_i = 1|s_j = r_i = 1) = 1$. If $s_j = r_i = 0$, then $\Pr(\hat{\pi}_i = 1|s_j = r_i = 0) = 0$. If $s_j \neq r_i$ and the value of s_j is observed, then $\Pr(\hat{\pi}_i = 1|s_j, s_j \neq r_i, s_j \neq x) = 1 - s_j$. If $s_j \neq r_i$, and the value of s_j is unobserved, then $i \in \mathcal{R}$ chooses an action $\pi_i = Y$, where

$$Y = \Pr(\hat{\pi}_i = 1|s_j \neq r_i, s_j \neq x, \theta, \varphi) = \frac{\Pr(s_j(\hat{\pi}_j) \neq 1|\cdot)\theta}{\Pr(s_j(\hat{\pi}_j) \neq 1|\varphi)\theta + \Pr(s_j(\hat{\pi}_j) \neq 0|\varphi)(1-\theta)}. \quad (1)$$

For all speakers $i \in \mathcal{S}$, i's expected utility is given by

$$\begin{aligned} E[u_i(\pi, \hat{\pi}_i)|s_i, s_{-i}(\hat{\pi}_{-i}); \varphi, \theta, q] &= -\sum_{k \in \mathcal{R}} [\Pr(r_k = 1 \in \{s_i, s_{-i}\}|s_i, s_{-i}(\hat{\pi}_{-i}), \cdot)(1 - \hat{\pi}_i)^2 \\ &+ \Pr(r_k = 0 \in \{s_i, s_{-i}\}|s_i, s_{-i}(\hat{\pi}_{-i}), \cdot)\hat{\pi}_i{}^2 \\ &+ \Pr(r_k \neq s' \in \{s_i, s_{-i}\}\backslash\{x\}|s_i, s_{-i}(\hat{\pi}_{-i}), \cdot)(q(1 - s' - \hat{\pi}_i)^2 + (1-q)(Y - \hat{\pi}_i)^2) \\ &+ \Pr(s_i = s_{-i} = x|s_i, s_{-i}(\hat{\pi}_{-i}), \cdot)(\theta - \hat{\pi}_i)^2]. \end{aligned}$$

Note that this expression is a function of q directly and of φ via Y. Before proceeding with the characterization of equilibria under different debate protocols, it is instructive to consider the relationship between the incentives created for the speakers by changes in the parameters q and φ. As q increases, receivers have greater abilities to infer their own types from unpersuasive speech, regardless of the speaker's type. As such, higher q creates a greater disincentive for speakers of either type to speak. Although φ also affects incentives to speak, it does so in a way that affects different types of speakers differently. As φ increases, a receiver will choose lower values of the action choice if she is not persuaded by speech (and cannot observe the speech directly). Thus, speaking is more appealing for speakers of type 0, but less appealing for speakers of type 1, for higher φ. An increase in q mitigates the importance of this effect of φ, since higher values of q imply that the receiver is less likely to have to choose her action in ignorance of her type. Further, in addition to the effect on receivers' action choices, changes in φ affect speakers' assessments of the likelihood of different possible speeches from the other speaker, which in turn affects their own choices whether to speak.

Our first result is the following instrumental lemma, which establishes the monotonicity of speaking choices with respect to the parameters q and φ, given that the other speaker's behavior (and, in the sequential case, the speaker's behavior at other times in the debate) is to be silent (choose x):

Lemma 1 *Suppose at all $t' \neq t$, both speakers choose silence, and at t, speaker i chooses silence as well. Then, for all $q \geq q_0^* = \theta$, for all φ, the best response at t of speaker j of type 0 ($\hat{\pi}_j = 0$) is also silence; for all $q < q_0^*$, there exists $\varphi_0^*(q) \in (0,1)$ such that the best response at t of that speaker is silence for all $\varphi < \varphi_0^*(q)$ and speech $s_j^t = 0$ for all $\varphi > \varphi_0^*(q)$. Similarly, for all $q \geq q_1^* = 1 - \theta$, for all φ, the best response at t of speaker of type 1 ($\hat{\pi}_j = 1$) is silence; for all $q < q_1^* = 1 - \theta$, there is $\varphi_1^*(q) \in (0,1)$ such that that speaker's best response at t is $s_j^t = 1$ for all $\varphi < \varphi_1^*(q)$ and silence for all $\varphi > \varphi_1^*(q)$.*

This lemma holds regardless of the rules of debate. The characterization of equilibrium behavior is contingent on the rules of debate.

Simultaneous Speech

Under simultaneous speech, a speaker i's pure strategy s_i is a mapping $s_i : \{0,1\} \to \{\hat{\pi}_i, x\}$, where the domain of the mapping is the set of i's possible true types and the range is the set of possible arguments. The following proposition characterizes the set of equilibria in debates with simultaneous speech protocols.

Proposition 1 *1. For all φ, whenever $q < 1$, there exists an equilibrium in which both types of speakers choose to speak, i.e., $s_j(\hat{\pi}_j) = \hat{\pi}_j$ for all $\hat{\pi}_j \in \{0,1\}$.*

2. There exists an equilibrium in which both types of speakers are silent, i.e., $s_j(\hat{\pi}_j) = x$ for all $\hat{\pi}_j \in \{0,1\}$, as well as a mixed-strategy equilibrium in which both players speak with a non-degenerate probability, if and only if either

(a) $q \geq \max\{\theta, 1-\theta\}$; or

(b) $q \leq \min\{\theta, 1-\theta\}$ *and* $\varphi_1^*(q) \leq \varphi \leq \varphi_0^*(q)$; *or*

(c) $\theta \leq q \leq 1-\theta$ *and* $\varphi \geq \varphi_1^*(q)$; *or*

(d) $1-\theta \leq q \leq \theta$ *and* $\varphi \leq \varphi_0^*(q)$.

The intuition for this proposition is as follows. Speaking is always a best response to speaking because, given that the other speaker speaks, one's own speech is either redundant (has no additional effect on the receivers' choices) or it serves to persuade all the receivers who were both unpersuaded by the other speaker's speech and unable to identify its content directly. Since these receivers must be of one's own type, persuading them results in their choosing actions identical to one's own ideal point. Silence is a best response to silence, however, only when each type of speaker prefers the receivers' learning nothing to hearing only her speech. In order for the speakers to have such an induced preference, it must be that receivers are able to infer a great deal about their type when they are unpersuaded by the speech - i.e., when their types do not match that of the speaker, and thus when their preferred actions are the opposite of those the speaker prefers. Such inferences are possible when q is sufficiently high or when the receivers place a sufficiently high probability on the speaker's being of the type that she actually is. In either case, the receivers will often enough adopt positions sufficiently far away from the speaker's ideal point when they are unpersuaded to make the speaker worse off in expectation if she speaks.

Note that the no-speech equilibrium is Pareto superior for the speakers both to the mixed-strategy equilibrium and to the pure-strategy equilibrium with speech.

Sequential Open-Ended Speech

Under sequential open-ended speech, a speaker i's pure strategy at some t such that $i \in O^t$, s_i^t, is a mapping $s_i^t : \{0,1\} \times \{0,1,x\}^{t-1} \rightarrow \{\hat{\pi}_i, x\}$. Note that, unlike under simultaneous speech, the domain of the mapping is a cross-product of the set of i's possible true types with the speech history prior to t.

Our result on debates with sequential open-ended speech offers a very resolute prediction with

respect to the equilibrium path of play. Its logic may remind one of the famous statement by Mark Twain: "It is by the goodness of God that in our country we have those three unspeakably precious things: freedom of speech, freedom of conscience, and the prudence never to practice either."

Proposition 2 *For all $q \in [0,1]$ and $\varphi \in [0,1]$, no arguments are revealed in equilibrium.*

The core intuition for this result in our model is the balance of counter-arguments. Under this debate rule, one always has a chance to "counter-argue" by choosing a speech with the opposite value. 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 "counter-arguing" - making the opposite argument. But, given preference concavity, such a counterargument would make the previous speaker worse off than she would have been had no one spoken at all, thus deterring her from making her argument. Note that, although there is no speech on the path of play, the balance of counter-arguments that sustains the no-argument equilibrium here relies on opposing speakers' ability to speak. 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.

Sequential Fixed-End Speech

Our last theoretical result suggests that the fixed-end debate rule can substantially change the incentives faced by the speakers relative to open-ended debate:

Proposition 3 *1. There exists a unique equilibrium in which both types of both speakers are silent if and only if either*

(a) $q \geq \max\{\theta, 1-\theta\}$; *or*

(b) $q \leq \min\{\theta, 1-\theta\}$ *and* $\varphi_1^*(q) \leq \varphi \leq \varphi_0^*(q)$; *or*

(c) $\theta \leq q \leq 1-\theta$ *and* $\varphi \geq \varphi_1^*(q)$; *or*

(d) $1-\theta \leq q \leq \theta$ *and* $\varphi \leq \varphi_0^*(q)$.

2. *There exists a unique equilibrium in which both types of both speakers speak if* $q \leq \min\{\theta, 1-\theta\}$ *and* $\varphi_0^*(q) \leq \varphi \leq \varphi_1^*(q)$.

3. *Under all other vectors of parameter values, in the unique equilibrium the speech behavior is contingent on the speaker types and the speaking order. The equilibrium outcome may be one in which both speakers speak, neither speaker speaks, or only the last speaker speaks.*

The first thing to note about this result is that it points to the possibility of asymmetric behavior among the speakers on the equilibrium path of play, which was ruled out in the equilibria of the other two debate rules. Whereas under those rules, both speakers either speak or not, here, it is possible that only one speaker may choose to speak. Second, a variety of behavior is possible on the path of play, because behavior is contingent on speaker type and order. Yet, there is always a unique equilibrium. This contrasts with the multiplicity of equilibria under simultaneous speech and the lack of contingency on speaker type and order under both simultaneous and sequential open-ended rules. In the end, the equilibria for this debate rule underscore the important role of the speaking order in affecting the informational content of debate and the overall ideological posture of the arguments aired in it. They suggest that we are likely to see more speech under this rule than under the open-ended sequential speech rule.

The contrasting properties of equilibria under sequential fixed-end speech hinge on the fact that, although a speaker always prefers to speak if the other speaker has spoken (or will speak with certainty), she may or may not prefer to speak if only one type of the other speaker would choose to speak after a history of silence. In particular, in part 3 of the proposition, one possible type of last speaker prefers to speak after a history of silence, while the other possible type prefers silence. It follows, then, that if the other speaker is of the former type, she will prefer silence. To see this, note that if the last speaker is in fact of her type and will speak even after a history of silence, then the receivers will hear the speech that she would have made even if she does not speak herself; in contrast, if the last speaker is in fact the opposite of her type and would be silent after a history of silence, then her speaking will result in the receivers' hearing both arguments instead of neither,

making the speaker worse off in expectation. However, if the first speaker is of the type that would be silent if she were last, then the first speaker will prefer to speak: if the last speaker is the same type, provoking her to speak does no harm, and if she is the opposite type, then the last speaker speaks regardless of the first speaker's behavior, and the first speaker is better off in expectation if the receivers hear her speech as well. In this fashion, the speech behavior observed in equilibrium depends on the actual types of the speakers and the order in which they may speak.

Robustness

In this section we consider the robustness of the equilibrium analysis presented above to generalizing our model in three immediate ways. First, we have assumed in the preceding that speakers are drawn from the same prior distribution of types. This implies that receivers do not have prior information about the speakers that could be used to distinguish them from one another, and that each speaker knows nothing more about his counterpart than he would about any randomly drawn speaker. Because these implications deviate from many settings of real-world debate, it is valuable to consider the equilibrium effects of allowing the agents to have different priors about speakers. In fact, the equilibria characterized above are robust to this possibility.

In the case of simultaneous speech, the existence of the equilibrium in which both speakers make potentially persuasive arguments depends neither on φ nor on φ being the same for different speakers. With respect to the existence of an equilibrium in which both speakers are silent (regardless of their types), observe that receivers' having more accurate beliefs about a speaker's true type makes speaking less advantageous for the speaker; if her speech is unpersuasive to a given receiver, this receiver will then choose an action farther from that speaker's ideal point. Thus, such equilibria continue to exist if receivers have speaker-specific beliefs, though over a different range of parameters.

To see that the unique equilibrium in the case of sequential open-ended debate is also robust to having speaker-specific beliefs, observe that, regardless of these beliefs, speech is the best response to speech. If there is a high probability that the other speaker is of the opposite type, then concavity

of preferences insures that she prefers silence to speech by both. If there is a high probability that the other speaker is of the same type, then, since speakers and receivers share common beliefs, receivers are likely to infer that they are the opposite type if they are unpersuaded by either speech, and again, concavity of preferences insures that the speaker prefers silence in expectation. In the case of sequential fixed-end debate, the partition of the parameter space corresponding to different types of equilibrium behavior changes, but the equilibria themselves are unchanged. In particular, notice that the equilibrium in which asymmetric speaking behavior is possible is robust in that there are still conditions under which only one type of last speaker would prefer speech to silence after a history of silence, and the logic of the remainder of the speaking strategies is unaffected by φ.

Our second consideration regarding robustness concerns the effects of letting each speaker j choose speech from the set $\{0, 1, x\}$ rather than $\{s_j = \hat{\pi}_j, x\}$:

Remark 1 *The equilibria described above are robust to letting $s_j \in \{0, 1, x\}$ for all $j \in \mathcal{S}$.*

Thus, modeling speaker choice from a larger set $\{0, 1, x\}$ leads to the same results as those described above for the model with $s_j \in \{\hat{\pi}_j, x\}$.

Finally, we also note that our results are robust to relaxing our assumption that agents update their beliefs about optimal choices efficiently (that is, in accordance with Bayes Rule). In an experimental study of deliberation, Dickson, Hafer, and Landa (2008) describe the existence of agents who, in violating *negative introspection*, fail to make inferences from unpersuasive arguments in settings where a fully Bayesian agent would be able to do so. These agents also fail to foresee that *others* might make such inferences from unpersuasive arguments and, accordingly, as senders, they "overspeak." Suppose that, in our model, there was a non-degenerate probability that a given receiver, but not a sender, was an agent of this type. If that probability is sufficiently high, then, holding fixed the response from the other debater, there is no downside to speech. However, the other debater now has a dominant strategy to speak, and the expected policy outcome of speech becomes the same as it is in the present model with fully Bayesian agents. The effect

under the sequential open-ended speech rule would be to reproduce the equilibrium path of play in Proposition 2, while under the simultaneous and sequential fixed-end rules, there will always be a unique equilibrium under which both speakers speak - behavior we showed to be in equilibrium for some parameter values under these rules when all agents are Bayesian. When the probability of receivers' violating negative introspection is perceived to be sufficiently low, the equilibrium predictions revert precisely to those described in Propositions 1 and 3.

A speaker who herself violates negative introspection will fail to appreciate the possible downside to speaking, holding fixed the other speaker's response at silence. If the probability that speakers violate negative introspection is sufficiently high, then under the simultaneous speech rule, there is, once again, a unique equilibrium with both debaters speaking. Under sequential open-ended speech, the downside to speaking comes from the expectation of counter-argument by the other speaker, a strategic insight which is independent of negative introspection, and so the equilibrium prediction is consistent with the characterization in Proposition 2. Under sequential fixed-end speech, the parameter range under which the equilibrium prediction is affected is one that supports the equilibrium in which some type of the last speaker will prefer to be silent. If the last speaker is non-negatively introspective, she will prefer to speak regardless of her type. With a high enough prior probability that the last speaker is non-negatively introspective and so will speak, the other speaker will prefer to speak if the probability that the audience finds the unconvincing argument uninformative is high (if the latter probability is sufficiently low, then to provoke speech from the penultimate speaker, the probability that the last speaker is not negatively introspective may have to be higher still.)

V. Experiment

We conducted a laboratory experiment as a means of empirically evaluating the predictions of our theoretical model. Our experimental design allows us to test two key insights derived from the model – first, that details of debate rules can greatly influence the amount of information that

individuals may be willing to communicate during debate; and second, that the extent to which such rules *do* matter can vary as a function of the informational environment.

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. 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.[3] 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 the model described in Section III. In each round, subjects earned a number of

[3]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.

"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 "true number" (that is, ideal point) 0 (10) within a discrete action space $\{0, 1, 2, 3, ..., 10\}$.[4] The receiver's ideal point 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 matched a receiver's argument, the receiver was told what the speaker's original choice had been, for example, "10." If a given speech did not match a receiver's argument, 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 actor's ideal point).

[4]Thus, our experimental instantiation added an additional assumption to the model in Section III, namely that the set of speakers was commonly known to include one speaker of each type. Note that, because speakers know their own types, this implies that speakers know the type of the other speaker as well.

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 had been.

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. A 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.[5]

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, which we refer to as involving Informative Unpersuasive Speeches (IUS), 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, which we refer to as involving Uninformative Unpersuasive Speeches (UUS), it was not possible

[5]A literal instantiation of the Open-Ended Sequential Speech game form would have been susceptible to "hijacking" by a small number of subjects who could have chosen to continue sending the same message over and over again in perpetuity. To avoid this, our instantiation of this debate protocol differed from that described in Section III, though we stress that subjects faced the same strategic incentives to give speeches (or not) in our instantiation as do agents in our game-theoretic model of Open-Ended Sequential Speech. Briefly, 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, "empty message." Debate then terminated at the first point in time following consecutive choices of "empty message" by the two speakers. This instantiation retained the critical feature driving incentives in our theoretical analysis, namely that each speaker would *always* have an opportunity to respond to a speech by the other (if she had not previously given her own speech herself), thus deterring speech in the first place because of the concavity of speakers' utility functions. Note that this experimental protocol does *not* instantiate the Fixed-End Sequential debate rule; in particular, 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*. See the instructions for further information on the framing of our debate protocols to subjects.

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 environment by *not* informing Receivers of the type of speaker from whom each message had come.[6] This informational environment thus corresponds to the parameter values $q = 0$ and $\varphi = \frac{1}{2}$ in our model. In contrast, the informational environment in the IUS treatments can be thought of either as corresponding to a setting in which $q = 1$, or to a setting in which $\varphi \in \{0, 1\}$. In either case, unpersuasive messages are fully informative. We note that, while our theoretical results indicate the existence of multiple equilibria over a range of parameter regimes, our model makes unique equilibrium predictions for the parameter values we employ in our experiment.

* TABLE 1 ABOUT HERE *

Table 1 depicts these theoretical predictions about speaker behavior on the equilibrium path for each of our four treatments (as well as offering information about participation in our laboratory sessions). In our treatments involving Informative Unpersuasive Speeches (IUS), speakers are deterred from making speeches in equilibrium under *either* of our debate rules. In contrast, in our treatments involving Uninformative Unpersuasive Speeches (UUS), 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 debate environment derived from our theoretical model.

[6]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.

Experimental Results: Listeners

Because speakers' losses 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 when possible; no significant differences in listener behavior were observed across treatments for quantities which could be compared across treatments.

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" through speech from a speaker, they in fact do so 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). A mere 2.6% (15/577) of listeners' choices fell further away than 5 from their ideal point. This pattern is suggestive that subjects' deviations from our theoretical expectations are not driven primarily by misunderstanding of the experimental scenario. Indeed, subjects' responses to our post-experiment debriefing questionnaire strongly suggest that the bulk of such deviations were due to fairness concerns; for example, by choosing 3 instead of 0, a listener with ideal point 0 would increase the payoffs of one of the speakers by 51 tokens, at a cost of only 9 to herself (and to the other speaker).[7] Results were similar in the IUS treatments when listeners received only an informative unpersuasive speech (choosing their ideal point 77.1% of the time, 37/48).[8]

[7]For example, 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."

[8]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 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

The other expectation from our theoretical framework is 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 not uncommon. Here, the debriefing questionnaires offer some evidence that many of these deviations can be explained by some subjects' tastes for risk.

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 under Uninformative Unpersuasive Speeches (UUS)

When unpersuasive speeches are uninformative, our theoretical analysis suggests that each speaker will perceive an incentive to send her speech under a debate rule of Simultaneous Speech, but that each speaker will instead perceive an incentive to send no speech under an Open-Ended Sequential rule. Figure 1 offers a first glimpse into our data from the UUS treatments, plotting the fraction of the time that speakers chose to send a speech over the course of our experimental sessions. For the Open-Ended Sequential treatment, the data in the graph reflects only the first choice made by the first-moving speaker.[9]

* FIGURE 1 ABOUT HERE *

The data in the figure indicate a considerable across-treatment difference, in the direction

environment varies, and therefore need the ability exogenously to manipulate the extent to which subjects can make such inferences. In our framework, this exogenous variation is captured by the parameter q, which can be interpreted as a proxy for the cognitive complexity involved in making this inference in specific debate settings.

[9]We do this here because speakers' best responses at subsequent information sets are conditional on the history of play up to that point. Tables 2 and 3 will offer a more detailed glimpse into how play unfolds over the Open-Ended Sequential extensive form.

predicted by the theory. Averaging over all periods, speakers in the UUS-Simultaneous treatment chose to send a speech 67.1% of the time (456/680); in contrast, first-moving speakers in the UUS-Open Ended Sequential treatment chose to send a speech only 34.7% of the time (111/320). Notably, this across-treatment difference became increasingly pronounced as subjects gained experience over the course of the experimental sessions. UUS-Simultaneous speakers continued to send speeches at a relatively constant rate over time (first five periods: 68.8%; last five periods: 67.1%), while first-moving UUS-Open Ended Sequential speakers did so less and less often as time progressed (first five periods: 51.25%; last five periods: 20.0%). Consistent with visual intuition from the graphs, simple probit specifications do not indicate a significant time trend in the UUS-Simultaneous data ($z = 0.30$, $p = 0.766$) but do in the UUS-Open Ended Sequential data ($z = -4.74$, $p < 0.001$).

* TABLE 2 ABOUT HERE *

The data in Figure 1 describe all of the decisions made by speakers in the UUS-Simultaneous sessions, but, as noted above, describe only the first speaker decision made during each round in the UUS-Open Ended Sequential sessions. Naturally, speakers' best responses at subsequent information sets will depend on the prior history of play. Table 2 offers a more complete depiction of play in the UUS-Open Ended Sequential sessions. These data offer further evidence of convergence towards our theoretical expectations. Over the course of the UUS-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, broadly in line with the first-mover decisions reported above. 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 (successfully activated the listener's "latent argument"). 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)).

These results have been posed 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 would be necessary for her to learn her own type and so make a fully informed action choice. In the UUS-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 UUS-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: $z = -3.92$, $p < 0.001$). Strikingly, as subjects become more experienced with the experimental protocol, debate is informative *less* often in the UUS-Open Ended Sequential treatment than in the UUS-Simultaneous treatment, even though speakers have the opportunity to respond to one another's speech under the Open-Ended Sequential but *not* under the Simultaneous speech protocol. 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 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 UUS-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 1, speaker behavior exhibits no time trend in the UUS-Simultaneous sessions; 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. In the debriefing questionnaire, a substantial number of subjects report

that they refrained from making speeches out of concerns for fairness or ethics.[10] 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 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.

Experimental Results: Speakers under Informative Unpersuasive Speeches (IUS)

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 2 ABOUT HERE *

Figure 2 depicts data from our IUS treatments in a format completely analogous to Figure 1. First, note that the pronounced pattern of divergence that was evident in Figure 1 for UUS sessions is clearly absent from the IUS data in Figure 2. Second, averaging over all periods, speakers in the IUS-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 IUS-Open Ended Sequential treatment, who chose to send a speech 33.3% of the time (80/240). A simple probit specification indicates

[10] 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 and 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."

a significant decrease in the rate at which speakers send speeches both for the IUS-Simultaneous treatment ($z = -6.95$, $p < 0.001$) and for the IUS-Open Ended Sequential Treatment ($z = -2.76$, $p = 0.006$).

* TABLE 3 ABOUT HERE *

Table 3 depicts our IUS data in the same format as Table 2 did for our UUS sessions. Once again, the more detailed data offer further support for our theoretical predictions. Over the course of the IUS-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. This pattern conforms closely not only with the first-mover decisions from this treatment, but also with the comparable figures from the UUS-Open Ended Sequential treatment. As in the UUS-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 IUS-Open Ended Sequential and IUS-Simultaneous treatments, and that these rates decrease sharply over the course of experimental sessions.[11] In the IUS-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%). In the UUS-Open Ended Sequential treatment, listeners learned their type 55.4% (133/240) of the time, a figure that

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

also decreased sharply over the course of the sessions (first five periods: 75.0%; last five periods: 38.3%).

VI. Conclusion

When engaging in debate, how do agents choose when to communicate their "best arguments" (and when not to)? How are agents' choices affected by the environment in which debate takes place – if they are indeed affected at all?

This paper has addressed these questions from both theoretical and experimental perspectives. We developed a game-theoretic model of debate, in which speakers may attempt to influence the ultimate actions of audience members. We traced speakers' incentives to deliver speeches or to remain silent as key features of the institutional and informational environment are varied. Specifically, we derived speakers' equilibrium behavior in the context of three distinct debate rules. Holding each debate rule fixed, we also explore the effects of a key informational factor – the extent to which each given member of the audience is able to extract informational content from exposure to an argument that he or she finds unconvincing.

Our findings indicate that speakers' strategic incentives are strongly influenced both by institutional and informational factors. When the informational content of unconvincing arguments is relatively high, speakers' best interests are served by remaining silent, refraining from advancing their arguments at all. In contrast, when the informational content of unconvincing arguments is relatively low, 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 equilibrium analyses suggest that the informational content of debate may vary in striking ways across different deliberative environments. We put these predictions to the test in a laboratory experiment, in which debate rules and informational conditions are varied in the context of a two-by-two experimental design. We find 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 results.

These findings suggest that the selection of rules for debate may have important normative as well as positive consequences. During the course of a debate, audience members may hear arguments that allow them better to understand their own underlying interests, and ultimately, choose actions that come closer to those that would maximize their welfare. However, debate can play such a role only if it is actually informative to its audience. Our results indicate that the rules under which debate takes place, and the informational environment in which it takes place, can exert considerable influence over its ability to perform this function.

Appendix

Lemma 1

Proof $s_j^t = x$ yields higher expected utility than $s_j^t = \hat{\pi}_j$ for $\hat{\pi}_j = 0$ iff

$$-\left(\theta q + \theta(1-q)\left(\frac{(1-\varphi)\theta}{(1-\varphi)\theta + \varphi(1-\theta)}\right)^2\right) \leq -\theta^2$$

and $s_j^t = x$ yields higher expected utility than $s_j^t = \hat{\pi}_j$ for $\hat{\pi}_j = 1$ iff

$$-\left((1-\theta)q + (1-\theta)(1-q)\left(\frac{\varphi(1-\theta)}{(1-\varphi)\theta + \varphi(1-\theta)}\right)^2\right) \leq -(1-\theta)^2.$$

Manipulating these two inequalities, we obtain equivalent inequalities

$$q\left(1 - \left(\frac{(1-\varphi)\theta}{(1-\varphi)\theta + \varphi(1-\theta)}\right)^2\right) + \left(\left(\frac{(1-\varphi)\theta}{(1-\varphi)\theta + \varphi(1-\theta)}\right)^2 - \theta\right) \geq 0 \qquad (2)$$

$$q\left(1 - \left(\frac{(1-\theta)\varphi}{(1-\varphi)\theta + \varphi(1-\theta)}\right)^2\right) + \left(\left(\frac{(1-\theta)\varphi}{(1-\varphi)\theta + \varphi(1-\theta)}\right)^2 - (1-\theta)\right) \geq 0, \qquad (3)$$

respectively.

Consider condition (2) first. If $\varphi = 0$, then (2) is true $\forall q\ \forall\theta$. If $\varphi = 1$, then (2) is true iff $q \geq \theta$. The left-hand side of (2) is increasing in q, and because it is quadratic in φ, (2) holds at equality at at most two values of φ. Thus for $q < \theta$, there exists a unique $\varphi_0^*(q)$ s.t. $\forall\varphi < \varphi_0^*(q)$, $\hat{\pi}_i = 0$

prefers $s_j = x$ to $s_j = \hat{\pi}_i$ and $\forall \varphi > \varphi_0^*(q)$, $\hat{\pi}_j = 0$ prefers $s_j = \hat{\pi}_j$ to $s_j = x$ (given silence by others and at all other times).

Next let $q > \theta$. Solving for the quadratic roots φ s.t. (2) holds at equality, we obtain $\varphi_{1,2} = \frac{2\theta - q \pm \sqrt{(q-1)(q-\theta)}}{2\theta(4\theta-1)}$. By assumption, $q - \theta > 0$, but $q - 1 < 0$; thus, the discriminant is negative and there are no real values φ s.t. (2) holds at equality. Thus, $\forall q \geq \theta$, $\hat{\pi}_j = 0$ prefers $s_j = x$ to $s_j = 0$.

Consider next condition (3). If $\varphi = 1$, then (3) is true $\forall q$ $\forall \theta$. If $\varphi = 0$, then (3) is true iff $q \geq 1 - \theta$. The left-hand side of (3) is increasing in q, and because it is quadratic in φ, (3) can hold at equality at at most two values of φ. Thus for $q < 1 - \theta$, there exists a unique $\varphi_1^*(q)$ s.t. $\forall \varphi < \varphi_1^*(q)$, $\hat{\pi}_j = 1$ prefers $s_j = 1$ to $s_j = x$ and $\forall \varphi > \varphi_1^*(q)$, $\hat{\pi}_j = 0$ prefers $s_j = x$ to $s_j = 1$.

Now let $q > 1 - \theta$. Solving for the quadratic roots φ s.t. (3) holds at equality, we obtain $\varphi_{1,2} = \frac{(2\theta-1)(q+\theta-1) \pm (1-\theta)\sqrt{(q-1)(q+\theta-1)}}{(3\theta-2)(q+\theta-1)+(1-\theta)^2}$. Given that $q + \theta - 1 > 0$ and $q - 1 < 0$, there are no real values φ s.t. (3) holds at equality. Thus, $\forall q \geq 1 - \theta$, $\hat{\pi}_j = 1$ prefers $s_j = x$ to $s_j = 1$. ■

Proposition 1

Proof

1. Let $q < 1$. If $\hat{\pi}_{-i} = \hat{\pi}_i$, then i obtains the same utility from $s_i = \hat{\pi}_i$ and $s_i = x$, given $s_{-i} = \hat{\pi}_{-i} = \hat{\pi}_i$. If $\hat{\pi}_{-i} = 1 - \hat{\pi}_i$, then in expectation, proportion $(1 - q)$ of $\{j : j \in \mathcal{R}$ and $\hat{\pi}_j \neq \hat{\pi}_{-i}\}$ choose $\pi_j = \hat{\pi}_i$ if $s_i = \hat{\pi}_i$ and $\pi_j = \frac{(1-\varphi)\theta}{(1-\varphi)\theta + \varphi(1-\theta)}$ is $s_i = x$; the action choices of other receivers are unaffected by s_i given $s_{-i} = \hat{\pi}_{-i}$. Thus, $s_i = \hat{\pi}_i$ is a best response to $s_{-i} = \hat{\pi}_{-i}$ and a unique best response if $q < 1$, $\varphi \neq 1 - \hat{\pi}_i$. (Note that if $q = 1$, then speech is ruled out by action weak dominance, given concavity of speakers' preferences.)

2. The proof of Lemma establishes the existence of the critical values φ_0^* and φ_1^* such that the no-speaking equilibrium exists for the regions identified in the Proposition. Next we show that there exists no equilibrium s.t. for some $\hat{\pi}_i' \in \{0, 1\}$, $s_i(\hat{\pi}_i') = \hat{\pi}_i'$ and for $\hat{\pi}_i'' \neq \hat{\pi}_i'$, $s_i(\hat{\pi}_i'') = x$. Suppose such a strategy profile. Then any $j \in \mathcal{R}$ has posterior beliefs $\Pr(\hat{\pi}_j = \hat{\pi}_i' | \hat{\pi}_j \neq s_i) = 0$; thus if $\hat{\pi}_j = s_i(\hat{\pi}_i') = \hat{\pi}_i'$, $\pi_j = \hat{\pi}_i'$, and if $\hat{\pi}_j \neq s(\hat{\pi}_i') = \hat{\pi}_i'$, $\pi_j = \hat{\pi}_i''$. But the concavity of i's preferences over π_j implies that type $\hat{\pi}_i'$ prefers $s_i = x$ to $s_i = \hat{\pi}_i'$ - a contradiction. By the

same logic, there exists no equilibrium in which one type plays a pure strategy and the other type plays a non-degenerate mixed strategy. Thus, in any mixed-strategy equilibrium, both types play non-degenerate mixed strategies.

To construct the mixed-strategy equilibrium, let, with some abuse of notation, $\sigma_1 = \Pr(s_i = 1|\hat{\pi}_i = 1)$ and $\sigma_0 = \Pr(s_i = 0|\hat{\pi}_i = 0)$. Each type $\hat{\pi}_i \in \{0, 1\}$ must be indifferent in expectation between $s_i = \hat{\pi}_i$ and $s_i = x$, given the other player's strategy (σ_0, σ_1). Let $U_i(s_i, s_{-i}, \hat{\pi}_i)$ be i's indirect expected utility as a function of speech and her type (suppressing φ, θ, and q for notational convenience). Then the conditions of indifference for $\hat{\pi}_i = 1$ and for $\hat{\pi}_i = 0$ are

$$\begin{aligned} & \varphi(1-\sigma_1)U_i(x,x,1) + (1-\varphi)[\sigma_0 U_i(x,0,1) + (1-\sigma_0)U_i(x,x,1)] \\ = \;& \varphi(1-\sigma_1)U_i(1,x,1) + (1-\varphi)[\sigma_0 U_i(1,0,1) + (1-\sigma_0)U_i(1,x,1)] \end{aligned}$$

and

$$\begin{aligned} & \varphi[\sigma_1 U_i(x,1,0) + (1-\sigma_1)U_i(x,x,0)] + (1-\varphi)(1-\sigma_0)U_i(x,x,0) \\ = \;& \varphi[\sigma_1 U_i(0,1,0) + (1-\sigma_1)U_i(0,x,0)] + (1-\varphi)(1-\sigma_0)U_i(0,x,0). \end{aligned}$$

Let

$$Z \equiv \frac{U_i(x,x,1) - U_i(1,x,1)}{U_i(1,0,1) - U_i(x,0,1) + U_i(x,x,1) - U_i(1,x,1)}$$

and recall that $U_i(1,x,1) = U_i(x,1,1) = U_i(1,1,1)$ and $U_i(0,x,0) = U_i(x,0,0) = U_i(0,0,0)$. Then, solving for σ_0 and σ_1, we obtain

$$\begin{aligned} \sigma_0 &= \frac{Z}{1-\varphi}\left[1 - \frac{[U_i(x,x,0) - U_i(0,x,0)](1-Z)}{[U_i(x,x,0) - U_i(0,x,0)](1-Z) + U_i(0,1,0) - U_i(x,1,0)}\right] \\ \sigma_1 &= \frac{[U_i(x,x,0) - U_i(0,x,0)](1-Z)}{\varphi[[U_i(x,x,0) - U_i(0,x,0)](1-Z) + U_i(0,1,0) - U_i(x,1,0)]}. \end{aligned}$$

■

Proposition 2

Proof We adopt the notational convention that $O^1 = \{1\}$ and $O^2 = \{2\}$. Recall that speakers' types are their private information, so knowing the speaking order does not, by itself, communicate information about the speakers' types.

First, we establish by contradiction that there is no equilibrium in which exactly one type of speaker will be the first to speak, i.e., in which for some $\hat{\pi}_i \in \{0,1\}$, $s_1^1(\hat{\pi}_i) = \hat{\pi}_i$ and $s_1^1(1-\hat{\pi}_i) = x$ or $s_2^2(\hat{\pi}_i) = \hat{\pi}_i$ and $s_2^2(1-\hat{\pi}_i) = x$ given $s_1^1 = x$. Suppose an equilibrium with these features. Then each receiver is able to infer with certainty her own type after hearing the first argument. But concavity of speaker preferences implies that the speaker prefers in expectation no speech - a contradiction. Thus, in any equilibrium, either both types of speaker choose to speak before another speaker or both make arguments only after another speaker has already spoken.

For each possible type of speaker 2, $\hat{\pi}_2 \in \{0,1\}$, if $q < 1$ then $s_2^2 = \hat{\pi}_2$ is a best response to $s_1^1 = \hat{\pi}_1$: if $\hat{\pi}_2 = \hat{\pi}_1$, then the outcomes of $s_2^2 = \hat{\pi}_2$ and $s_2^2 = x$ are the same; but if $\hat{\pi}_2 \neq \hat{\pi}_1$, then $s_2^2 = \hat{\pi}_2$ persuades receivers who were unpersuaded by $\hat{\pi}_1$ to adopt $\pi = \hat{\pi}_2$, increasing 2's utility. (If $\forall i \in \mathcal{R}\ r_i = \hat{\pi}_i$, then $s_2^2 = \hat{\pi}_2$ and $s_2^2 = x$ produce the same outcomes, and are thus, bot best responses. Similarly, if $q = 1$, speaker 2 is indifferent over $s_2^2 = \hat{\pi}_2$ and $s_2^2 = x$. In all other cases, $s_2^2 = \hat{\pi}_2$ is the unique best response.)

Given that $s_2^2 = \hat{\pi}_2$ if $s_1^1 = \hat{\pi}_1$ and (by the same logic) $s_1^3 = \hat{\pi}_1$ if $s_2^2 = \hat{\pi}_2$ and $s_1^1 = x$, an equilibrium in which $s_1^1(\hat{\pi}_1) = x$ and $s_2^2(\hat{\pi}_2, s_1^1 = x) = x\ \forall \hat{\pi}_1\ \forall \hat{\pi}_2$ is supported iff $\forall \hat{\pi}_i \in \{0,1\}$, $\hat{\pi}_i$ prefers in expectation $s_1^1 = s_2^2 = x$ to $s_1^1 = \hat{\pi}_1$ and $s_2^2 = \hat{\pi}_2$:

$$-\varphi\left[(1-\theta)q + (1-\theta)(1-q)\left(\frac{\varphi(1-\theta)}{\varphi(1-\theta)+(1-\varphi)\theta}\right)^2\right] - (1-\varphi)(1-\theta) < -(1-\theta)^2$$

and

$$-(1-\varphi)\left[\theta q + \theta(1-q)\left(\frac{(1-\varphi)\theta}{\varphi(1-\theta)+(1-\varphi)\theta}\right)^2\right] - \varphi\theta < -\theta^2.$$

Simplifying and re-arranging terms, we obtain

$$(q-1)\varphi[(1-\varphi)^2\theta^2 + 2\varphi(1-\varphi)\theta(1-\theta)] + \theta[(1-\varphi)\theta + \varphi(1-\theta)]^2 > 0$$

and

$$(1-\varphi)[(1-\varphi)^2\theta^2 + q(2\varphi(1-\varphi)\theta(1-\theta) + \varphi^2(1-\theta)^2] + (\varphi-\theta)[(1-\varphi)\theta + \varphi(1-\theta)]^2 > 0.$$

Observe first that the left-hand side of each inequality is increasing in q, and that each inequality is true if $q = 1$. Next, observe that at $q = 0$, both inequalities are true $\forall \varphi \in [0,1]\ \forall \theta \in [0,1]$. Thus,

both inequalities are true $\forall\varphi\ \forall\theta\ \forall q$, and $\forall\varphi\ \forall\theta\ \forall q$ there exists an equilibrium in which $s_1^1(\hat{\pi}_1) = x$ $\forall\hat{\pi}_1$ and $s_2^2(\hat{\pi}_2, x) = x\ \forall\hat{\pi}_2$.

Lastly, observe that although $\hat{\pi}_1$ is indifferent between $s_1^1 = x$ and $s_1^1 = \hat{\pi}_1$ if $s_2^2 = \hat{\pi}_2$ (given $s_1^3(\hat{\pi}_1, (s_1^1, \hat{\pi}_2))$), $\hat{\pi}_1$ strictly prefers in expectation $s_1^1 = x$ to $s_1^1 = \hat{\pi}_1$ if $s_2^2 = x$ with positive probability, i.e., $s_1^1 = x$ weakly action dominates $s_1^1 = \hat{\pi}_1$. Similarly, after $s_1^1 = x$, 2 prefers $s_2^2 = x$ to $s_2^2 = \hat{\pi}_2$, given $s_1^3(\hat{\pi}_1, (x, \hat{\pi}_2)) = \hat{\pi}_1$. ■

Proposition 3

Proof

1. Lemma 1 establishes that the given conditions are necessary and sufficient for both types of the last speaker ($j \in O^o$) to prefer x to $\hat{\pi}_j$ in response to silence from all previous speakers. Each type of each speaker i prefers $\hat{\pi}_i$ to x in response to speech from any other speaker: if she is the same type as the previous speaker, her speech has no effect, and if she is the other type, her speech only causes some of the receivers who were unpersuaded by the previous speech to adopt positions closer to her own. As established in the proof of Proposition 2, all types of all speakers prefer in expectation no speaker speaking to both speakers speaking; thus all types of all speakers choose x under the given conditions.

2. Lemma 1 establishes that the given conditions are necessary and sufficient for both types of the last speaker ($j \in O^o$) to prefer $\hat{\pi}_j$ to x in response to silence from all previous speakers; thus under these conditions, $\hat{\pi}_j$ is a dominant strategy for $\hat{\pi}_j \in \{0, 1\}$, $j \in O^o$. Anticipating that $s_j^o(\hat{\pi}_j, \cdot) = \hat{\pi}_j$, speaker $i \neq j$ prefers $s_i^t(\hat{\pi}_i, \cdot) = \hat{\pi}_i$ at some $t < o$ such that $i \in O^t$: if $\hat{\pi}_i = \hat{\pi}_j$, then the outcome is unaffected by s_i^t, but if $\hat{\pi}_i \neq \hat{\pi}_j$, then i is strictly better off choosing $s_i^t(\hat{\pi}_i, \cdot) = \hat{\pi}_i$ over $s_i^t(\hat{\pi}_i, \cdot) = x$.

3. In the remaining cases, exactly one type of the last speaker ($j \in O^o$) prefers $\hat{\pi}_j$ to x in response to silence from all previous speakers, and the other type prefers x to $\hat{\pi}_j$. Consider the subcase in which type 0 prefers speech 0 and type 1 prefers x; the argument for the symmetric subcase is the same. Both types choose $\hat{\pi}_j$ in response to speech from another speaker.

Consider then the best response of each possible type of speaker $i \neq j$. Suppose $\hat{\pi}_i = 1$. Then if for some $t < o$, $s_i^t = 1$, the receivers hear 1 if $\hat{\pi}_j = 1$ and both 0 and 1 if $\hat{\pi}_j = 0$; and if $s_i^t = x$, the receivers hear no speech if $\hat{\pi}_j = 1$ and 0 if $\hat{\pi}_j = 0$. Thus the best response for $\hat{\pi}_i = 1$ is $s_i^t = 1$. Suppose $\hat{\pi}_i = 0$. Then if for some $t < o$, $s_i^t = 0$,the receivers hear both 0 and 1 if $\hat{\pi}_j = 1$ and 0 if $\hat{\pi}_j = 0$; and if $s_i^t = x$, the receivers hear no speech if $\hat{\pi}_j = 1$ and 0 if $\hat{\pi}_j = 0$. Thus the best response for $\hat{\pi}_i = 0$ is $s_i^t = x$.

Combining the best responses of i and j, either i and j both speak (if $\hat{\pi}_i = 1$); i and j are both silent (if $\hat{\pi}_i = 0$ and $\hat{\pi}_j = 1$); or i is silent and j speaks (if $\hat{\pi}_i = 0$ and $\hat{\pi}_j = 0$).

■

Remark 1

Proof First observe that for $\varphi \in (0,1)$, $q < 1$, any speaker i prefers in expectation receivers' hearing both $\hat{\pi}_i$ and $(1 - \hat{\pi}_i)$ to hearing $(1 - \hat{\pi}_i)$ alone. Thus, i prefers in expectation $s_i = \hat{\pi}_i$ to $s_i = 1 - \hat{\pi}_i$ given $s_{-i} \neq x$. It remains to show that if receivers anticipate $s_i(\hat{\pi}_i) = \hat{\pi}_i$, then i, in fact, prefers $s_i = \hat{\pi}_i$ to $s_i = 1 - \hat{\pi}_i$ if $s_{-i} = x$.

Recall that $\Pr(\hat{\pi}_i = 1 | s_j \neq r_i, s_j \neq x, \theta, \varphi) = Y$ as given in (1) above. If $s_j = 1$, then $\Pr(\pi_i = 1|\cdot) = \theta$; $\Pr(\pi_i = 0|\cdot) = (1-\theta)q$; $\Pr(\pi_i = Y|\cdot) = (1-\theta)(1-q)$. If $s_j = 0$, then $\Pr(\pi_i = 1|\cdot) = \theta q$; $\Pr(\pi_i = 0|\cdot) = (1-\theta)$; $\Pr(\pi_i = Y|\cdot) = \theta(1-q)$. If $s_j = x$, $\Pr(\pi_i = \theta|\cdot) = 1$.

Suppose $s_j \neq r_i$ and let $s_j(\hat{\pi}_j) = \hat{\pi}_j$. Consider first $\hat{\pi}_j = 1$.

If $s_j = 1$, then j's expected utility, given that j is the only speaker, is

$$-\sum_{i \in \mathcal{R}} \left[(0)\theta + (1)(1-\theta)q + \left(\frac{\varphi(1-\theta)}{(1-\varphi)\theta + \varphi(1-\theta)} \right)^2 (1-\theta)(1-q) \right]. \tag{4}$$

If $s_j = 0$, then it is

$$-\sum_{i \in \mathcal{R}} \left[(0)\theta q + (1)(1-\theta) + \left(\frac{\varphi(1-\theta)}{(1-\varphi)\theta + \varphi(1-\theta)} \right)^2 \theta(1-q) \right]. \tag{5}$$

Given that $s_j(\hat{\pi}_j) = \hat{\pi}_j$ and $\hat{\pi}_j = 1$, $s_j = 1$ is better than $s_j = 0$ if quantity (4) is greater than (5). That condition is certainly satisfied if $\forall j \in \mathcal{R}$:

$$(1-2\theta) \left(\frac{\varphi(1-\theta)}{(1-\varphi)\theta + \varphi(1-\theta)} \right)^2 < (1-\theta),$$

which is always true $\forall\theta > 0$.

Consider next $\hat{\pi}_j = 0$. If $s_j = 1$, then j's expected utility, given that j is the only speaker, is

$$-\sum_{i\in\mathcal{R}}\left[(1)\theta + (0)(1-\theta)q + \left(\frac{(1-\varphi)\theta}{(1-\varphi)\theta + \varphi(1-\theta)}\right)^2(1-\theta)(1-q)\right].$$

If $s_j = 0$, then it is

$$-\sum_{i\in\mathcal{R}}\left[(1)\theta q + (0)(1-\theta) + \left(\frac{(1-\varphi)\theta}{(1-\varphi)\theta + \varphi(1-\theta)}\right)^2\theta(1-q)\right].$$

Comparing these expressions, $s_j = 0$ is certainly better than $s_j = 1$ if $\forall j \in \mathcal{R}$:

$$(2\theta - 1)\left(\frac{(1-\varphi)\theta}{(1-\varphi)\theta + \varphi(1-\theta)}\right)^2 < \theta,$$

which is always true $\forall\theta < 1$.

It follows that if $s_j(\hat{\pi}_j) = \hat{\pi}_j$, then neither type will deviate to $s_j = 1 - \hat{\pi}_j$. ■

References

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

[2] Austen-Smith, David, and Timothy Feddersen, 2006. "Deliberation, Preference Uncertainty, and Voting Rules." *American Political Science Review* 100 (2), 209-217.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

[20] Meirowitz, Adam. 2006. "Designing Institutions to Aggregate Preferences and Information." *Quarterly Journal of Political Science* 1 (4), 373-392.

[21] Mendelberg, Tali. 2002. "The Deliberative Citizen: Theory and Evidence." *Political Decision Making, Deliberation and Participation* 6, 151-93.

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

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

[24] Patty, John. "Arguments-Based Collective Choice." *Journal of Theoretical Politics*, forthcoming.

Table 1. The Experimental Sessions: Participation and Theoretical Expectations

Treatment	Nr of Sessions	Nr of Subjects	Predicted Equilibrium Path
UUS-Open Ended Sequential	3	48 (=21+15+12)	Speakers Always Send x ("Empty Message")
UUS-Simultaneous	3	51 (=21+18+12)	Speakers Always Make Speech ("Send True Number")
IUS-Open Ended Sequential	2	36 (=18+18)	Speakers Always Send x ("Empty Message")
IUS-Simultaneous	2	30 (=18+12)	Speakers Always Send x ("Empty Message")

Table 2. Frequencies With Which Speakers Made Speeches: UUS Treatments

	Simultaneous	O.E. Sequential (S1, T1)	O.E. Sequential (S2, T1) after S1 silent	O.E. Sequential (S2, T1) after S1 unpersuasive	O.E. Sequential (S2, T1) after S1 persuasive	O.E. Sequential (S1, T2) after S2 unpersuasive	O.E. Sequential (S1, T2) after S2 persuasive
all periods	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)
periods 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)
periods 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)
periods 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)
periods 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.)

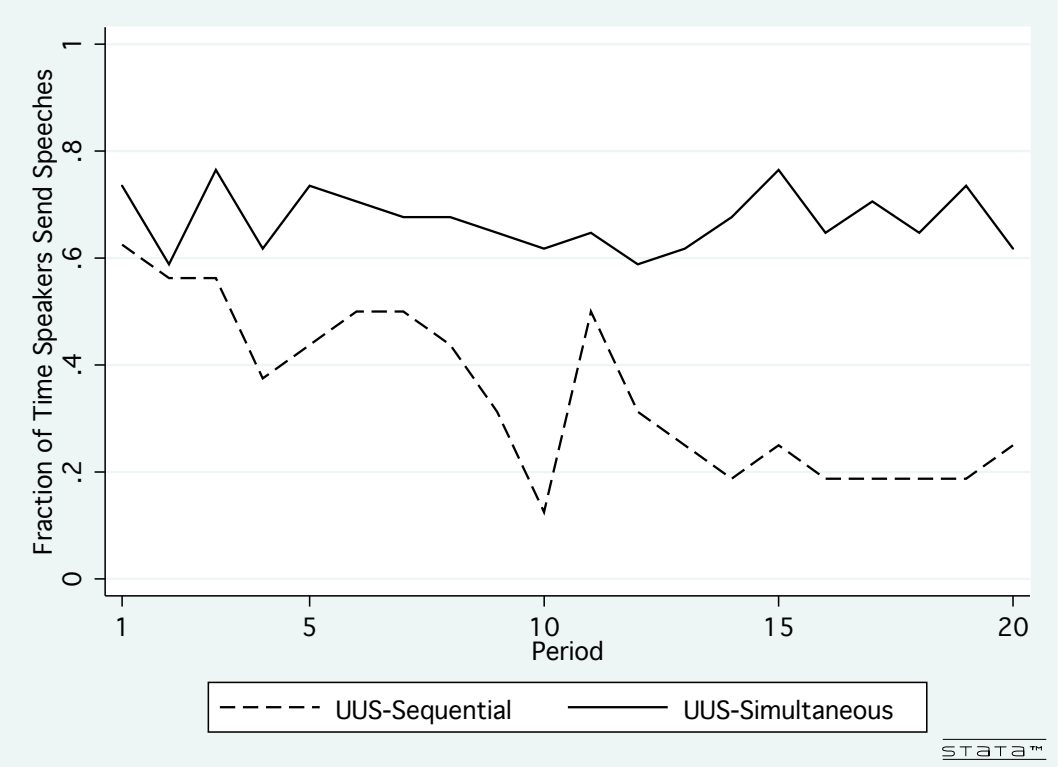
Note: O.E. Sequential = Open Ended Sequential, S1 = 1st Speaker, S2 = 2nd Speaker, T1 = 1st Turn, T2 = 2nd Turn

Table 3. Frequencies With Which Speakers Made Speeches: IUS Treatments

	Simultaneous	O.E. Sequential (S1, T1)	O.E. Sequential (S2, T1) after S1 silent	O.E. Sequential (S2, T1) after S1 unpersuasive	O.E. Sequential (S2, T1) after S1 persuasive	O.E. Sequential (S1, T2) after S2 unpersuasive	O.E. Sequential (S1, T2) after S2 persuasive
all periods	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)
periods 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)
periods 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)
periods 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)
periods 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.)

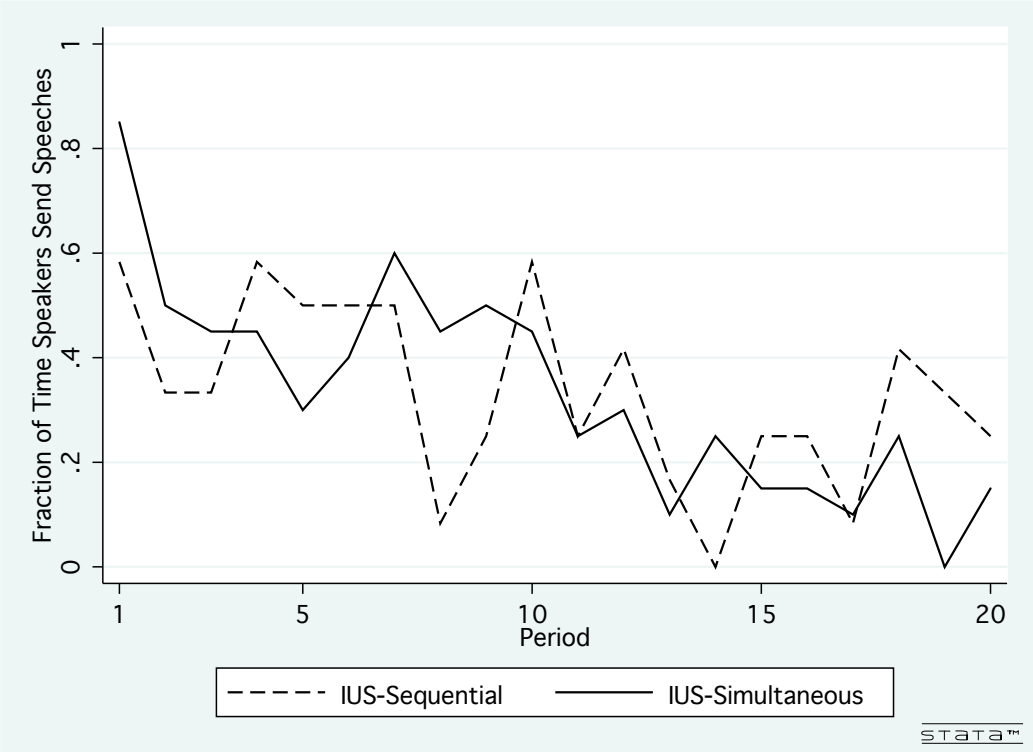
Note: O.E. Sequential = Open Ended Sequential, S1 = 1st Speaker, S2 = 2nd Speaker, T1 = 1st Turn, T2 = 2nd Turn

Figure 1: Frequencies With Which Speakers Made Speeches in the UUS Treatments (by Round).



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

Figure 2: Frequencies With Which Speakers Made Speeches in the IUS Treatments (by Round).



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