Rules of Debate*

Catherine Hafer†           Dimitri Landa‡

03/08/2006

Preliminary - Comments Welcome

Abstract

We analyze the effects of debate rules on the informational quality of debates. We show that these effects differ fundamentally across debate rules and are importantly conditioned on the nature of participants’ cognition and learning and on their mode of policy-making (private vs. public politics).

I. Introduction

Moral and political debates on the choice of a policy or the justifiability of an action are typically envisioned as contests of opponents marshalling their best, most persuasive arguments in an attempt to sway the audience in favor their preferred alternatives and against the competition. However, with a diverse audience, if arguments are persuasive to some, they are likely to be unpersuasive to others, and the very fact of their potential persuasiveness carries with it a danger that the unpersuaded will turn against the alternative favored by the speaker: after all, they are receiving evidence that the reasons that appear to underlie the appeal of that alternative do not - to them - hold water. This suggests that, contrary to the popular image, making one’s best arguments in the debate may not always be the best course of action. What should we, then, expect from public debate? Under what institutional and political circumstances will individuals with diverse interests and reasons share their reasons with each other? Put differently, when will the public debate be informative and improve the quality of decision-making?

*Prepared for the Princeton University Conference on Deliberative Democracy.

†Assist. Prof. of Politics, NYU, e-mail: catherine.hafer@nyu.edu

‡Assist. Prof. of Politics, NYU, e-mail: dimitri.landa@nyu.edu
In contemporary democratic theory, the free exchange of arguments and the responsiveness of individual and collective decision-making to debate have emerged as the new standard of political legitimacy. Although the mechanisms supporting this view vary, most of them associate the positive effect of deliberation with its promotion of the quality of individual and/or collective decision-making. To the extent that post-deliberative decision-making reflects the information or the “best arguments” communicated in the course of deliberation, the resulting decisions must be seen as meeting a higher standard of normative acceptability.

The expectation of successful, informative communication that underlies these arguments cannot, however, be taken for granted, and the need to better understand the determinants of such communication is increasingly attracting the attention of game theorists and political economists. The present paper contributes to this understanding by analyzing the effects of adopting different rules of debate on the informational content of debate in different political environments.

Debate as the marshalling of reasons or arguments for and against the alternatives on the table may often be best construed as a communication in which speakers attempt to explain to their audiences the perceived logical consequences of their beliefs. For instance, a fundamentalist Christian might point out to another fundamentalist that, if they believe every word of the Bible is true, then they must believe that life begins at conception, and thus must be pro-life. A person who is not persuaded by such an argument in debate tends not so much to contest the credibility or veracity of the speaker as her foundational assumptions (e.g., that every word of the Bible is true) or the arguments she has used to connect those assumptions to the desired conclusion. Thus, much of what is going on in such debates is not the transmission of new information about facts or states of the world, but the teasing out of arguments. The speakers in them are convincing because, for the relevant listeners, their arguments are intrinsically persuasive (they are sound and valid), rather than because they themselves are perceived as credible sources of information.

Our model of debate focuses on such intrinsically persuasive argumentation aimed at a heterogenous audience. Different agents find different reasons or arguments compelling, and these reasons can affect their beliefs about and choices of policy. For each agent, some of these compelling reasons are active: she knows that these reasons are both valid and relevant, and they determine her perceived optimal choice. Other of these compelling reasons are latent: she is initially unsure of either their validity or relevance, but will be convinced of both when she hears arguments that correspond to them. Hearing such an argument “activates” the corresponding latent reason, turning it into an active one.

We show that different rules of debate create different strategic incentives for the participants. These incentives depend, further, on the mode of policy-making (majoritarian “public politics,” in which policy decisions are collectively determined through the aggregation of preferences by majority rule, and “private politics,” in which policy decisions are individual and affect others through utility externalities. They also depend on the speakers’ beliefs about the cognitive capacities of their audience (in particular, their beliefs about the proportion of Bayesian agents, for whom even invalid arguments may be informative, as opposed to those agents who confound validity and relevance and so fail to make the possible policy-relevant inferences from messages that do not correspond to either their active or latent reasons). The informational consequences of debate are contingent on those incentives and suggest the importance of tailoring the rules of debate to the political circumstances.
framing the policy choice.

Three of our results deserve special emphasis. First, public politics creates particularly strong incentives for the participants in debate to make their potentially informative arguments. Second, sequential speech with endogenous consensual debate termination, which arguably comes closest to the deliberative democratic ideal, leads in the context of private politics to the least informative debate. Third, under plausible assumptions on behavioral agency, simultaneous speech in the context of private politics produces maximally informative 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 model, including the formal definitions of the informational environment, different rules of debate, and the cognitive agencies of the audience. In Section IV we present the equilibrium analysis of the model. Section V concludes with a brief recap. The appendix at the end of the paper contains formal proofs of the supplementary results invoked in the text.

II. Relation to the Existing Literature

The role of deliberation in democracies more generally is the focus of a vast literature. Some of the key works in it include Manin (1987); Cohen (1996); Gutmann and Thompson (1996); and Bohman and Rehg (1997). The existing formal literature on deliberation focuses, by and large, on the analysis of cheap-talk models of information transmission. In these models, individuals exchange messages about the content of their private information, and because that information is private, they can, within limits, misrepresent their information. The determination of the extent or quality of deliberation in such models, then, turns largely on the credibility of the sent messages, which is induced by the proximity of the sender’s primitive preferences to those of the individual receiver (Crawford and Sobel 1982) or of the expected majorities in the group of receivers (Meirovitz 2003, 2005), and by the sender’s pivotalness both as a source of information and as a voter (Austen-Smith and Feddersen 2005, 2005b). Austen-Smith and Feddersen (2005b) and Gerardi and Yariv (2002) consider the effects of preference aggregation rules on the informativeness of speech in the cheap-talk environment.

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 (2005), 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. By contrast, in the model we analyze, messages are fully provable and complete, yet their veridicality (truth content) differs across agents. The most natural examples of this setting are precisely the messages in which arguments are of the “self-discovery” nature described in the Introduction.

Calvert and Johnson (1998) emphasize the prominence of arguments free of private information in motivating their view of deliberation as a mechanism for coordinating expectations in the context of multiple distributionally distinct equilibria. This approach, like that of Patty (2005), in whose model arguments affect collective decision-making but not individual
preferences over alternatives, is also consistent with rational agency. But while deliberation may serve the functions motivating these models, arguments free of private information can often also be seen as persuading, rather than only coordinating, and appreciation of the causes and consequences of their persuading properties is fundamental for understanding the phenomenon of public deliberation. Aragones et al. (2001) make the case that arguments by analogy can have such properties without relying on the revelation of private information and consider the analytical structure of inductive updating in response to them. The closest models to the one developed in the present paper are Glazer and Rubinstein (2001) and Hafer and Landa (2003 and 2006). The latter analyze related models with deliberation that, like in the present model, satisfies full provability and private veridicality. They focus on the welfare properties of the equilibria of a game in which agents choose how to allocate their deliberative resources between receiving (and processing) arguments and making arguments to others and on individual choices of deliberative venues, respectively.

Glazer and Rubinstein consider the informational properties of debate rules, including simultaneous speech and the one-shot 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. Although they restrict speakers to making truthful arguments, their setup is closer to the standard cheap-talk model. Austen-Smith (1993) and Krishna and Morgan (2001) analyze the effects of different communications rules within a cheap talk model in which two speakers attempt to influence the actions by a third party. Battaglini (2002) examines a cheap-talk model with multiple senders and a multidimensional state variable. Ottaviani and Sorensen (2001) analyze the effects of orders of speech within a cheap-talk model of committee debate with speakers who have heterogeneous expertise and career concerns.

There exists a considerable experimental and behavioral literature on deliberation. Mendelberg (2002) provides a comprehensive review of that work up to that date. The deviations from Bayesian agency, the effects of which we explore below, receive direct experimental support in Dickson, Hafer, and Landa (2005) whose experiment is set in a strategic deliberative context closely related to the one in this paper. More indirect evidence, from decision-theoretic contexts, is supplied by the experimental studies of confirmatory bias (see Lord et al. 1979; Tetlock 1992; Dawes 1998; Rabin 1998; Baron 1994) and especially hypothesis testing (Wason 1968, 1977; Baron 1994, Ch. 13).

III. The Model

The Informational Framework

In order to operationalize \( i \)'s uncertainty over her ideal policy in a manner consistent with the notion of deliberation discussed above, suppose that \( i \)'s beliefs about her ideal policy choice are a function of the reasons she finds convincing. For each individual (or agent) \( i \), the list of such convincing reasons is \( r_i \in \{0, 1\}^n \), where the \( j \)th element of that list, \( r_{ij} \), \( j = 1, ..., n \), is an independent draw of a Bernoulli random variable with \( \Pr(r_{ij} = 1) = \theta_i \in [0, 1] \). The value of \( r_{ij} \) can be interpreted as signaling how convinced a given person is by opposing
arguments with respect to dimension \( j \), with \( r_i^j = 1 \) interpreted as \( i \) being convinced by a “right” argument and \( r_i^j = 0 \) as \( i \) being convinced by a “left” argument with respect to that dimension. (Because the value of \( \theta_i \) is agent-specific and each agent’s list of reasons \( r_i \) is independently drawn, different members of a society will not find the same arguments compelling.) The vector \( r_i \) can be thought of as \( i \)’s type and \( \theta_i \) as \( i \)’s summary ideological characteristic, determining how likely \( i \) is to be convinced by “right” (or, complementarily, “left”) arguments. We suppose that each \( i \) is uncertain about \( \theta_i \), but that the distribution of \( \theta_i \), described by the probability density function \( p(\theta) \), is common knowledge.

Nature randomly selects and reveals some dimensions of \( r_i \). Let \( \mathcal{A}_i^j \subseteq \{1, 2, \ldots, n\} \) contain the dimensions of the reasons that Nature reveals so that \( j \in \mathcal{A}_i^j \) if and only if \( i \) knows \( r_i^j \). Call \( r_i^j \) active if \( j \in \mathcal{A}_i^j \), and latent if \( j \notin \mathcal{A}_i^j \) and let \( \mathcal{L}_i^j \) be the set of all such latent dimensions, so that \((\mathcal{A}_i^j, \mathcal{L}_i^j)\) is a partition of the set of dimensions \( \{1, 2, \ldots, n\} \). Let \( l_i^j \) be the number of those revealed dimensions on which \( r_i^j = 1 \) and let \( m_i^j \) be the number of those revealed dimensions on which \( r_i^j = 0 \), so that \(|\mathcal{A}_i^j| = l_i^j + m_i^j\).

When referring to arguments \( r_i^j \), we identify them by their distinct theses, or distinct “argument labels,” for example, “the human fetus is a human being.” Since an argument typically consists of a series of premises, inferences, and conclusions, including the summary argument thesis/label, knowing the label is not the same as knowing, or being persuaded, or even necessarily knowing that one would be persuaded, by the argument for which it stands (as is the case in this example). However, the identification of an argument with a given label often enables agents to assess the probability of encountering the corresponding arguments in their social interactions as well as to determine whether a given thesis, and so the arguments supporting it are likely to comport with other arguments they know to be true. To simplify notation, we refer to both the argument itself and its label as \( r_i^j \), with the understanding that, in a general case, agents may, prior to deliberation, know only a label, and not whether they are convinced by its corresponding argument (unless the reason expressed by that argument is already active). Whether a heard argument \( s_i^j \) is one’s latent argument or not depends on whether one would accept the set of premises it employs. Thus, when referring to \( r_i^j \) as \( i \)’s latent reason, we mean that there is a set of premises that \( i \) accepts that would, on examination, imply to \( i \) the validity of the thesis \( r_i^j \). Finally, we assume that when presented with an argument, agents know what kind of argument it is - i.e., what label it has.\(^1\)

**Actions and Choices**

For simplicity, we assume that the population is divided into potential speakers and potential listeners. Let \( S \) be the set of senders/speakers and \( R \) the set of receivers/listeners. For each agent \( i \), we assume that \( \theta_i \in [0, 1] \), which we described as \( i \)’s summary ideological

\(^1\)This last assumption is, arguably, most constraining in the case in which the argument heard is unconvincing and unconvincing arguments are potentially informative. In that case, it may matter for how one updates whether that argument is believed to be the best argument for a particular thesis or not. Our identification of arguments with labels means, in effect, that agents who are in a position to make inferences from unconvincing arguments (agents whom we call B-agents below) can tell when an argument is the best argument for a given thesis and when it is, in fact, uninformative.
characteristic, is also i’s ideal point in the underlying policy space.

A player i’s (true) type $\theta_i$ is not observable, even to herself. To keep things simple, we assume that $\theta_i \sim U[0,1]$, i.e., that player i’s $\theta_i$ is randomly drawn from the uniform distribution on interval $[0,1]$. This fact is commonly known to both speakers and listeners. We assume, moreover, that speakers observe the initial biases $l’, m’$ of each of the receivers and of the other speakers.\footnote{This assumption, as opposed to, say, that the speakers know only the summary ideological parameters, $E[\theta_j]$ for the listener $j$, ensures that the speaker’ optimal choices are well-defined, but otherwise has no effect on the substantive results we present below.}

Recall that $l_i’$ is the number of dimensions randomly revealed by Nature at the beginning of play on which $r_j’i = 1$ and $m_i’$ is the number of those revealed dimensions on which $r_j’i = 0$. As a matter of interpretation, we suppose that potential speakers’ reasons are active, i.e., that $A_i = \{1, 2, \ldots, n\}$ and thus $l_i’ + m_i’ = n \forall i \in S$. The speakers’ ability to make potentially persuasive arguments on every issue dimension follows naturally from such a supposition. However, in the technical sense it plays no role in establishing the results that follow and thus is, strictly speaking, unnecessary.

We delay a precise description of the timing and frequency of the speakers’ speeches until the introduction of the different debate rules we consider. For each speaker $i$, a single speech $s_i$ consists of an ordered list of $n$ messages, corresponding to $n$ dimensions of $r$. Formally, $s \in \{0, 1, x\}^n$, where $x$ can be interpreted as uninformative or irrelevant remark, e.g., “this is a great country.” In what follows, we reserve the term “argument” only for those statements that are potentially persuasive (that is, informative because provable as right or wrong). Thus, a statement $x$ would not be referred to as an argument.

Thus, the speaker may choose to make a fully uninformative speech (choosing $s = (x, x, ..., x)$), or one that is potentially persuasive on all or some dimensions. Designate the set of dimensions of $s_i$ that consist of arguments as $D_i$, i.e., $D_i = \{d : s_i^d \in \{0, 1\}\}$ and let $s_i$ be determined as follows: at each decision node at which the speaker $i$ speaks, let $i$ choose a speaking behavioral strategy $(D_i, \sigma_i)$ conditional on the debate history. (If $i$ speaks at time $t$, we will also add a subscript $t$ and write $D_{it}$ to denote the set of dimensions on which speech is generated by $i$’s choice of $\sigma_i$ at $t$; $\sigma_{it}$.) Then $i$’s speech on dimension $d \in D_i$, $s_i^d$, is an independent draw of a Bernoulli random variable with $\Pr(s_i^d = 1) = \sigma_i$. For $d \notin D_i$, $s_i^d = x$.\footnote{Characterizing the behavioral strategies in this manner insures equilibrium existence in pure strategies, but otherwise has no effect on the substantive results we present below.}

Following the end of the debate, receivers make their policy choices in accordance with their updated beliefs and the given policy-making procedure.

**Debate Rules**

The game form is defined by a combination of the mode of policy selection, which we discuss in detail below, and the debate protocol. A debate protocol is a triple consisting of a set of speakers $S$, a set of speaker messages $\{0, 1, x\}^n$, and a debate rule. A debate rule is defined by a pair of a speaking order and a debate termination protocol.

A speaking order $O$ is an ordered list of speakers, $o$ elements long, such that $o \geq |S|$ and
∀i ∈ S, there is j ∈ {1, ..., o} s.t. the jth element of the list O, Oj, is i. Thus, the definition of the speaking order requires that each member of 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 (Simultaneous Speech)** The debate rule is **simultaneous speech** if the debate consists of and ends with one-time simultaneous speeches by all members of S.

Simultaneous speech debate rule 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 (Open-ended Sequential Speech)** The debate rule is **open-ended sequential speech** if (1) each speaker i has an opportunity to make a speech si 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 the speakers unanimously prefer to end debate.

Formally this debate rule can be described as follows. Given a speaking order O, at time t = 1, speaker O1 speaks and then ∀i ∈ S, i votes to close the debate or not. If the vote to close debate is unanimous, debate ends; otherwise debate continues. As long as debate continues, speaker Ot speaks in t, followed by a vote to close debate or not. If t ∈ {o + 1, ..., 2o}, the speaker in t is Ot−o; if t ∈ {qo + 1, ..., (q + 1)o} where q ∈ {1, 2, ...}, then the speaker in t is Ot−qo. Debate ends immediately upon a unanimous vote to close debate, after which the listeners select policy. We assume that given any two debates with identical sets of arguments made, the speakers prefer the debate that takes less time (this assumption allows us to avoid indeterminate length debates without altering the substance of any of the results).

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 (Fixed-end Sequential Speech)** The debate rule is **fixed-end sequential speech** if (1) each speaker i has an opportunity to make a speech si 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 O1 speaks, at t = 2, O2 speaks, etc. After Oo speaks, debate ends and the listeners select policy.

Both types of sequential speech debate rules are particularly common in assemblies. Note, however, that their definitions do not isolate unique objects: they 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.
Public vs. Private Politics

As noted in the introduction, we consider debate in relation to two different policy-making environments. The first of these environments is one of majoritarian preference aggregation. To keep things simple, we assume a fixed binary agenda, \( \{ \pi^1, \pi^2 \} \), with \( \pi^k \in [0, 1] \). Let \( \pi_\mu \) be the majority choice. Given this choice, we assume that \( i \)'s utility is captured by the quadratic loss function \( -(\pi_\mu - \theta_i)^2 \), which is the standard means of capturing risk aversion in the spatial models of policy choice. The substantive implications of our results extend to any concave loss function.

In contrast to the above setting of “public politics,” in “private politics” (Baron 2003), the individual policy choices of citizens are sovereign, but they may be influenced through the use of the public or political sphere.\(^4\) To model this policy-making environment, we assume that each agent’s payoffs depend directly on the policy choices of the members of the society as well as her own. We can write \( i \)'s utility as \( -(\pi_i - \theta_i)^2 - \eta \sum_{j \in \mathcal{R} \setminus \{i\}} (\pi_j - \theta_i)^2 \), where \( \eta \) is a measure of the importance of others’ choices relative to \( i \)'s own. Because the receiver cannot affect the policy choices of others, and hence her utility from their choices, her behavior is assumed to maximize her expected utility from her own choice, \( -\int_0^1 p(\theta|l'_i, m'_i)(\pi_i - \theta)^2 d\theta \).

Because speakers do not obtain new information in the course of deliberation, and hence their beliefs and the consequent policy choices remain constant with respect to their deliberative choices, each speaker \( i \)'s behavior is assumed to maximize her expected utility from the choices of the receivers, \( -\int_0^1 p(\theta|l'_i, m'_i)(\sum_{j \in \mathcal{R}} (\pi_j - \theta)^2) d\theta \).

Belief Updating and Cognition

The knowledge of \( r_i \) does not imply certain knowledge of \( \theta_i \). Accordingly, whether \( i \) is a speaker or a receiver, pre-deliberative beliefs about her own type \( \theta_i \) can be characterized by the probability density function

\[
p(\theta|l', m') = \frac{\theta^m(1 - \theta)^{m'}}{\int_0^1 \theta^m(1 - \theta)^{m'} d\theta}.
\]

From the common knowledge distribution of \( \theta_i \), players derive common conditional probabilities \( \Pr(r_i|\mathcal{A}'_i) \). No new information about the primitive probabilities \( \Pr(r_i) \) is available from playing the game. Although \( r_i \) cannot change, \( \mathcal{A}'_i \) can. We analyze two types of behavioral agency that are defined in relation to this change. Efficient updating upon receiving a message would require agents to update their beliefs about their reasons using a Bayes Rule. Updating this way has two implications: first, if the message received is one’s latent

\(^4\)Examples of issues which would most immediately fall into the domain of private politics include the desirability of obtaining abortions (holding constant the official policy); the work-force participation of women; the value of post-secondary education; choices concerning child-bearing, etc. More generally, our formulations of individual preferences could be thought of as the first-order approximation of preferences in the context of public politics, i.e., politics in which a single binding decision is made following deliberation.
argument, then the receiver should add that argument dimension to the set of her active dimensions; second, if the message received on a given dimension is neither one’s latent nor active argument but not an uninformative argument \( x \), then the receiver ought to infer what her latent reason on that dimension must be.\(^5\)

Call agents who update their policy positions in accordance with Bayes Rule at each information node in the history of play B-agents. Let \( s \) be the vector of revealed reasons. Then, for B-agents, the change in \( A_i \) follows the trajectory

\[
A_i = \{ j : j \in A_i' \text{ or } r_i^j = s^j \}.
\]

A natural and commonly observed deviation from the belief-updating expressed by this trajectory entails failing to make an inference from an unconvincing (invalid) argument (that is, nonetheless, provable). If agent \( i \) who updates this way receives an argument \( s^j \) that matches her type \( r_i^j \), it becomes (or stays) active; if \( s^j \) does not match \( r_i^j \), then she does not recognize the argument (that is, \( i \) learns \( r_i^j \) if and only if \( r_i^j = s^j \) and \( r_i^j \) was latent) and no change in \( A_i' \) occurs. In what follows we refer to such agents as NB-agents. Their learning can be summarized formally as

\[
A_i = \{ j : j \in A_i' \text{ or } r_i^j = s^j \}.
\]

The basic ontology of learning for the NB-agents is, then, that of recognizing latent reasons - reasons that agents are endowed with and would be able to embrace as “their own” after recognizing their fit with other held beliefs, but that are not actively available to them prior to deliberation for developing the corresponding policy position. In effect, they act as if they do not know that the information available to them has implications for what latent arguments they must or are likely to agree with, and so will fail to update their beliefs about their induced ideal policies accordingly. One can interpret this behavioral assumption as saying that agents change their policy positions only when they can give or understand a sound and valid propositional support for the newly adopted positions, at which point they switch to the policy implied by the conjunction of that (previously latent) reason and their initial active arguments. The “sound and valid propositional support” is a valid argument that proceeds from the premises that are shared by the listener. The listener’s “latent” reason may be thought of as precisely that argument.\(^6\)

We assume that speakers are B-agents themselves, understand how both types of agents update, and have a common prior belief that the proportion of B-agents in the audience is \( b \).

Before moving to the discussion of the equilibrium properties in this model, we state the

---

\(^5\)Let \( k \in \mathcal{L}_j \), then

\[
Pr(r_j^k = 0 | s_i^k = 1 \text{ and } r_j^k \neq s_i^k) = \frac{Pr(r_j^k \neq s_i^k | r_j^k = 0 \text{ and } s_i^k = 1) Pr(r_j^k = 0 | s_i^k = 1)}{Pr(r_j^k \neq s_i^k | r_j^k = 0 \text{ and } s_i^k = 1) Pr(r_j^k = 0 | s_i^k = 1)}.
\]

Because \( Pr(r_j^k \neq s_i^k | r_j^k = 1 \text{ and } s_i^k = 1) = 0 \), it follows that \( Pr(r_j^k = 0 | s_i^k = 1 \text{ and } r_j^k \neq s_i^k) = 1 \).

\(^6\)As we show elsewhere (Hafer and Landa 2005), NB-agents systematically and exclusively fail the condition of negative introspection - they do not know what they do not know. We show that their deviation from the Bayesian inference is, in fact, reducible to their violation of negative introspection.
following intuitive result that describes the responsiveness of listeners’ policy positions to speeches made in the debate, which, in our model, is independent of the rules of debate and the mode of policy-making.

**Lemma 1** (1) For any given speech \( \sigma \), an NB-listener’s policy position will, in expectation, move in the direction of that speech’s bias regardless of the listener’s prior position and remain unchanged if the speech is unbiased; (2) For any given speech \( \sigma \), a B-listener’s policy position will, in expectation, remain unchanged.

**Proof** See Appendix. ■

The intuition for this result is rather straightforward. If a speaker is more likely to produce right- than left-biased arguments (i.e., if \( \sigma > \frac{1}{2} \)), then a disproportionately high number of the NB-listener’s right-biased latent reasons will be activated. Thus, whatever that listener’s initial bias, she will, in expectation, move to the right in response to a right-biased speaker. Given that the listeners’ prior beliefs are correct and the speakers do not possess additional information on the likelihood that any given listener is of a particular type, the fact that the B-listeners learn from both potentially persuasive messages (0 and 1) means that the B-agents’ positions are, in expectation, unaffected by speech.

**IV. Equilibrium Analysis**

Throughout, our equilibrium concept 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 in the manners described above, depending on their cognitive types B or NB. The equilibrium concept is, thus, analogous to the Perfect Bayesian equilibrium, with the modification that NB-agents update their beliefs as described above rather than via Bayes’ Rule.

The following are the requirements this equilibrium concept imposes in the various environments we analyze:

*In simultaneous speech debates:*

Each speaker \( i \) chooses \( (D_i, \sigma_i) \) to maximize her expected utility given the strategies of the other speakers \( \{(D_j, \sigma_j)\}_{j \in \mathcal{R} \setminus \{i\}} \), the listeners’ prior beliefs and the nature of their belief updating, and the listeners’ policy selection strategies.

*In sequential speech debates:*

Let \( l_jt \) be the number of 1’s and \( m_jt \) the number of 0’s that \( j \in \mathcal{R} \) learns from the first \( t \) speeches. Let the debate history prior to time \( t \), be \( H_t \). \( H_t = (H_t^1, H_t^2, ..., H_t^n) \) where \( H_t^k \subseteq \{0, 1, x\} \) contains every argument made on dimension \( k \) in the first \( (t - 1) \) speeches. The behavioral strategy of time \( t \) speaker in the speaking order \( O \), \( O^t \), conditional on \( H_t \), is \( (D_{O^t}(H_t), \sigma_{O^t}(H_t)) \). She chooses this strategy to maximize her expected utility given \( H_t \), the strategies of subsequent speakers \( \{(D_{O^t'}(H_{t'}), \sigma_{O^t'}(H_{t'}))\}_{t' = t+1, ...} \), the listeners’ beliefs and the nature of their belief updating, and the listeners’ policy selection strategies. Further, in the open-ended sequential speech, at each time \( t \), \( \forall i \in S \), \( i \) votes to keep debate open if her expected utility from continuing the debate is higher than her expected utility from closing it immediately.

*In public politics:*
Assume, without loss of generality, that the voting agenda is \( \{\pi^1, \pi^2\} \) and that \( \pi^1 < \pi^2 \). Given the majoritarian model of public politics with a binary agenda and quadratic preferences, restriction of voting strategies to weak dominance requires that each listener \( j \in \mathcal{R} \) votes for the alternative that, if implemented, would give her the higher utility, given her beliefs after the debate. Let \( l_{jT} \) be the number of 1’s and \( m_{jT} \) be the number of 0’s that the listener \( j \) learns from the entire debate. The equilibrium requires, then, that \( \forall j \in \mathcal{R}, j \) votes for \( \pi^1 \) if and only if

\[
E[\theta_j | p(\theta) | l_j', m_j' + m_{jT})] \leq \frac{\pi^1 + \pi^2}{2}.
\]

In private politics:

In equilibrium, each \( j \in \mathcal{R} \) chooses \( \pi^*_j \) to maximize her expected utility, given her beliefs after the debate:

\[
\pi^*_j = E[\theta_j | p(\theta) | l_j', m_j' + m_{jT})] \quad \forall j \in \mathcal{R}.
\]

In characterizing the equilibria of the debate games we consider, we assume that the set of speakers is sufficiently diverse to include speakers who could, in principle, benefit from engaging in informationally consequential speech to move the audience to the right as well as those who would want to move it to the left. To give a more precise formulation of what we mean by this, say that a player is biased if her preferred policy position (or speech if indicated) is not \( \frac{1}{2} \). We refer to the player as being left-biased if it is less than \( \frac{1}{2} \) and right-biased if it is greater than \( \frac{1}{2} \). Similarly, if \( s^j = 0 \), we refer to the \( j \)th argument as left-biased, and if \( s^j = 1 \), as right-biased. We can now define our condition on the set of speakers:

**Condition 1 (Speaker Diversity)** For every set of listeners \( \mathcal{R} \), the set of speakers includes at least one speaker who, at \( t = 0 \), would be better off if every listener had one more active left-biased argument than if no listeners did, ceteris paribus, and at least one speaker who, at \( t = 0 \), would be better off if every listener had one more active right-biased argument than if no listeners did, ceteris paribus.

Formally, we require that given \( \mathcal{R} \), there exist \( i, h \in \mathcal{S} \) such that at \( t = 0 \),

\[
- \sum_{j \in \mathcal{R}} \int_0^1 p(\theta_i | l_i', m_i')(E[\theta_j | p(\theta) | l_j', m_j' + 1) - \theta_i)^2 d\theta_i > - \sum_{j \in \mathcal{R}} \int_0^1 p(\theta_i | l_i', m_i')(E[\theta_j | p(\theta) | l_j', m_j')] - \theta_i)^2 d\theta_i
\]

and

\[
- \sum_{j \in \mathcal{R}} \int_0^1 p(\theta_h | l_h', m_h')(E[\theta_j | p(\theta) | l_j', m_j' + 1) - \theta_h)^2 d\theta_h > - \sum_{j \in \mathcal{R}} \int_0^1 p(\theta_h | l_h', m_h')(E[\theta_j | p(\theta) | l_j', m_j')] - \theta_h)^2 d\theta_h.
\]

Three observations regarding this condition are worth noting. First, it does not imply that there are speakers who would prefer that any listener discover that she is persuaded by
one more right-(or left-)biased argument than she had thought. (It may be the case that, from the standpoint of each of those speakers, some listeners are already biased too much in the direction re-enforced by their speech.) Second, this condition does not ensure that informative speech will happen in equilibrium. Whether informative speech happens depends on elements of the deliberative environment that vary across the debate games we analyze (debate rules, private vs. public politics, and the expected proportion of the B-agents in the audience) and on the equilibrium that is being played in it. Third, this condition implies that, if one speaker succeeds in pulling the audience to the left, then another speaker must have an incentive to pull it back to the right and vice versa. Thus, at any point in the history of the debate, there is a speaker who, setting aside the strategic responses, would prefer to pull the audience toward herself.

Public Politics

Our main result regarding the effects of debate rules on the informational content of debates given public politics is as follows:

**Proposition 1** Regardless of the debate rule and the proportion of B-agents in the audience, in every debate equilibrium either all possible arguments are made or those arguments that are not made are, in equilibrium, inconsequential for the policy outcome.

**Proof** Without loss of generality, suppose a majority of voters prefers $\pi_1$ to $\pi_2$ before deliberation (at $t = 0$).

Simultaneous Speech:

Suppose that listener $j$ prefers $\pi_1$ and is a B-agent. The expected change in $E[\theta_j]$ from hearing speech generated by $\sigma_i = 1$ on $D \neq \emptyset$ is 0; but the probability that $E[\theta_j]$ increases enough for $j$ to choose $\pi_2$ over $\pi_1$ increases in $|D|$. Suppose $j$ is an NB-agent. Then, by Lemma 1, the expected change in $E[\theta_j]$ is positive and the probability that $j$ chooses $\pi_2$ over $\pi_1$ also increases in $|D|$. The effect of hearing $\sigma_i = 0$ is symmetric. Thus, for a speaker who prefers $\pi_2$, $\sigma_i = 1$ on $D = \{1, ..., n\}$ dominates $\sigma_i < 1$, $D \subseteq \{1, ..., n\}$ and $\sigma_i = 1$, $D \subset \{1, ..., n\}$. For the speaker who prefers $\pi_1$, the best response is to offer speech $\sigma = 0$ covering all dimensions, since B-agents are unaffected but, by Lemma 1, NB-agents will, in expectation, be weakly drawn to support $\pi_1$ instead of $\pi_2$.

Open-ended Sequential Speech:

Arguing as above, we obtain that for a time $t$ speaker $i$ who prefers $\pi_2$ if $\pi_1$ is favored by a majority, then $\sigma_{it} = 1$ on $D = \{1, ..., n\}$ dominates $\sigma_i < 1$, $D \subseteq \{1, ..., n\}$ and $\sigma_i = 1$, $D \subset \{1, ..., n\}$. If $i$ succeeds in obtaining majority support for $\pi_2$, then some speaker $h$ at time $t' > t$ who prefers $\pi_1$ chooses $\sigma_{ht'} = 0$ on $D_{ht'} \subseteq \{1, ..., n\}$ by the same argument; if $i$ fails, then further speech cannot affect the policy outcome, and thus the speakers are indifferent over all possible speech. The possible outcomes are:

1. $0 \in H^k_2 \forall k \in \{1, ..., n\}$ and $\pi_2$ is chosen by a majority; or
2. $1 \in H^k_2 \forall k \in \{1, ..., n\}$ and $\pi_1$ is chosen by a majority; or
3. $\{0, 1\} \in H^k_2 \forall k \in \{1, ..., n\}$.

When any one of these three conditions is met, no further speech can be consequential (i.e., can affect which policy is chosen), and thus all participants weakly favor ending debate.
Fixed-end Sequential Speech:

Suppose before the last speaker in $O$ speaks, the majority of voters prefers $\pi^1$ to $\pi^2$. If that speaker prefers $\pi^1$, she will make a speech of $(x, x, ..., x)$; if she prefers $\pi^2$ and no one has made a speech $s = (1, ..., 1)$, she will because the probability that voter $j$’s $E[\theta_j]$ increases enough for $j$ to choose $\pi^2$ over $\pi^1$ increases in $|D|$. If the last speaker prefers $\pi^1$, then given speaker diversity, at some point earlier in the order of speech there must be a speaker with an opposite preference and that speaker will prefer a speech $s = (1, ..., 1)$. If they succeed in swaying majority to $\pi^2$, then someone (last speaker, if necessary) will have a best response of $s = (0, ..., 0)$. If they fail, subsequent speakers are indifferent. Thus, as with open-ended speech, all arguments may be made, or just those contrary to the winning alternative. ■

It is worth emphasizing that the informational value of debate is fully induced by the arguments that are made by speakers whose policy preferences are, at the moment of speech, in opposition to the preferences of the majority. Moreover, when not all arguments are made, those made include all arguments that are against the chosen alternative.

Private Politics

We begin with a lemma that is instrumental in proving the main results in this subsection.

**Lemma 2** Given private politics, each speaker is worse off, in expectation, if (1) a B-agent hears an argument on a latent dimension; (2) an NB-agent hears both arguments on a latent dimension.

**Proof** See Appendix. ■

The intuition for this result is as follows. For every argument $s^j_k$ that a speaker $k$ makes, every listener $i$ who has the matching latent argument $r^j_i = s^j_k$ is convinced by it and updates her beliefs about policy accordingly. Every B-agent listener $i$ for whom $r^j_i$ is latent and $r^j_i \neq s^j_k$, must infer that she finds the opposite argument convincing—i.e. that her $j$th dimension argument is 1 if 0 was unconvincing and that it is 0 if 1 was unconvincing—and update her beliefs about policy accordingly. Because the listener is choosing her policy to maximize her expected utility given her beliefs, her policy choice (in the absence of further information) already strikes a balance between her possible true policy ideal points that takes into account the likelihood of her having each of them. Thus, the expected movement as a result of her learning another argument is 0. Similarly, the expected movement by an NB-agent who hears both arguments on a given dimension must also be 0. Because the speaker and each of these types of listener share the same ex ante beliefs about the listeners’ likely true policy ideals, the speaker has the same estimation of her listeners’ likely movement in response to additional information, but, because the speaker’s utility exhibits diminishing marginal returns in the closeness of the listeners’ policy choices, the speaker is harmed more by these listeners’ moving away than helped by their moving closer, ceteris paribus. This means that, in expectation, the speaker is worse off in both circumstances described in the lemma.
Simultaneous Speech

**Proposition 2** There exists a debate equilibrium in which all possible arguments are made if and only if there exists a non-zero probability that the audience includes NB-agents. If and only if the proportion of NB-agents is sufficiently small, there exists another equilibrium in which no arguments are made.

**Proof** Suppose \( b = 1 \). Then, by Lemma 2, if \( \forall j \in \mathcal{S}\backslash \{i\} \) \( s_j = (x, x, ..., x) \), then \( i \)'s best response is \( s_i = (x, x, ..., x) \). If \( \exists j \in \mathcal{S} \) s.t. \( D_j \neq \emptyset \), then \( D_i \subseteq D_j \) has no effect. Thus, \( s_i = (x, x, ..., x) \) is a weakly dominant strategy.

Suppose \( b < 1 \).

By Lemma 2, making arguments to B-agents is always costly. By Lemma 1, if NB-agents are exposed to biased speech, then, in expectation, their policy choices move in the direction of that bias (i.e., to the right if speech is right-biased and to the left if speech is left-biased). Thus, a sufficiently extreme speaker benefits from making arguments to NB-agents.

Suppose \( i \in \mathcal{S} \) and \( \forall j \in \mathcal{S}\backslash \{i\} \), \( s_j = (x, x, ..., x) \). Although the incidence of \( i \)'s costs and benefits from \( D_i \neq \emptyset \) depends on \( b \), their magnitudes do not. Hence, as \( b \) increases, making provable arguments becomes less desirable. While some moderates may prefer to send \( x \) even when \( b = 0 \), all speakers send \( x \) if \( b = 1 \). Thus, there exists \( b_i^* \in [0, 1) \) s.t. \( \forall b < b_i^*, D_i \neq \emptyset \) and \( \forall b > b_i^*, D_i = \emptyset \).

Suppose \( \exists j \in \mathcal{S} \) and \( \exists k \in \{1, ..., n\} \) s.t. \( s_j^k \in \{0, 1\} \). Then \( s_j^k \) affects only NB-agents. Thus \( k \in D_i \) if \( i \) is sufficiently extreme to benefit in expectation from attempting to persuade the NB-agents.

Consider two potential speakers, 1 and 2, such that 1’s optimal speech when she alone speaks is \( \hat{s}_1 < \frac{1}{2} \) and 2’s optimal speech when she alone speaks is \( \hat{s}_2 > \frac{1}{2} \). Assume \( D_1 = D_2 \neq \emptyset \). From Lemma 1, in expectation, the receivers will move left in response to \( \hat{s}_1 \) and right in response to \( \hat{s}_2 \). The optimality of \( \hat{s}_1 \) and \( \hat{s}_2 \) and the given strategy space implies that 1 wants the receivers to move left and 2 wants them to move right. If 1 is offering speech \( \hat{s}_1 \), then 2, in order to achieve her desired net move of the receivers to the right, must offer \( \sigma_2' > \hat{s}_2 \) and/or \( D_2 \supset D_1 \). Similarly, in the presence of 2’s speech, 1 must prefer a more extreme leftist speech \( \sigma_1' < \hat{s}_1 \) and/or \( D_1 \supset D_2 \). But then 2 must prefer \( \sigma_2 > \sigma_2' \), etc. Thus, in equilibrium, \( \sigma_1^* = 0, \sigma_2^* = 1, \) and \( D_1 = D_2 = (1, ..., n) \). Every receiver \( i \in \mathcal{R} \) learns \( r_i \), and by the same argument made in Lemma 2, every speaker obtains lower utility in expectation from attempting to persuade the members’ policy choices when they are informed than she does when they are uninformed.

If, for a given audience and given \( b \), there does not exist a potential speaker \( i \) s.t. \( b_i^* > b \), then for every potential speaker, making argument \( x \) is a best response to every other speaker’s making argument \( x \). Thus every speaker making argument \( x \) is an equilibrium. ■

Sequential Open-ended Speech

Our next result offers a very resolute prediction with respect to the equilibrium path of play in debates with sequential open-ended speech:

**Proposition 3** Regardless of the proportion of B-agents among the listeners, there is a unique equilibrium path of play on which no arguments are made.
Proof If \( b = 1 \), then the result follows immediately by Lemma 2. So suppose \( b < 1 \).

Suppose \( \forall i \in S, i \) prefers \( D_{it} = \emptyset \) to any \( D'_{it} \neq \emptyset \) \( \forall \sigma_{it} \in [0,1] \) if \( i \) anticipates that every other speaker \( j \) will choose \( D_{jt'} = \emptyset \) until debate ends. Then no arguments are made in equilibrium.

Suppose \( \exists i \in S \) and \( t \) s.t. if \( D_{jt'} = \emptyset \) \( \forall j \in S \{ i \} \) \( \forall t' > t \), then \( i \)'s optimal speech is \( D_{it} \) s.t. for some \( d \in D_{it}, H^j_{it} = \{ x \} \). Without loss of generality, suppose that \( \sigma_{it} < \frac{1}{2} \) and that \( E[\sigma_{it}] = E[\sigma_{jt}] \forall j \in S \). Observe that if \( i \) prefers \( (D_{it}, \sigma_{it}) \) s.t. \( D_{it} \ni d, \sigma_{it} < \frac{1}{2} \) to \( D_{it} = \emptyset \), then \( i \) also prefers \( D_{it} \) s.t. \( D_{it} = \{ d \} \) and \( \sigma_{it} = 0 \) to \( D_{it} = \emptyset \), thus, it is sufficient to establish that no sequentially rational strategy profile includes \( D_{it} \ni d \) where \( H^j_{it} = \{ x \} \).

Speaker Diversity implies that \( \exists j \in S, j \neq i \) s.t. \( \sigma_{it} = 0, D_{it} \ni d \), then at \( t' > t, j \) prefers \( \sigma_{jt'} = 1, D_{jt'} \ni d \) if \( j \) anticipates that \( \forall k \in S, t'' > t', D_{kt''} = \emptyset \). Let \( j \) be s.t. \( E[\sigma_{jt'}] = E[\sigma_{kt'}] \forall k \in S \).

Suppose \( i \) chooses \( \sigma_{it} = 0, D_{it} = \{ d \} \) after \( H^j_{it} = \{ x \} \) and, at \( t' > t, j \) chooses \( \sigma_{jt'} = 1, D_{jt'} = \{ d \} \). If, \( \forall t'' > t' \) assigned to \( i, D_{jt''} = \emptyset \), then \( \sigma_{jt'} = 1, D_{jt'} = \{ d \} \) is a best response for \( j \), but, from Lemma 2, \( i \) would have preferred \( D_{it} = \emptyset \). Thus, if \( \exists k \in D_{jt''} \) s.t. \( H^k_{it} = \{ x \} \) given \( H^j_{it} \ni \{ 0,1 \} \) and \( H^k_{it} = H^j_{it} \forall h \in \{ 1,\ldots,n \}\{d \} \), then in any sequentially rational action at \( H_t, \exists d \in D_{it} \) s.t. \( H^d_{it} = \{ x \} \).

If, after some number of rounds of \( i \) making new arguments and \( j \) making new counter-arguments, \( i \) will not want to make new arguments, then, from Lemma 2, \( i \) chooses not to make new arguments in any of her preceding speeches.

Suppose \( i \) has made an argument and \( j \) has made a counter-argument on \( n-1 \) dimensions, i.e. at time \( t, \exists d \in \{ 1,\ldots,n \} \) s.t. \( H^k_t \ni \{ 0,1 \} \forall k \in \{ 1,\ldots,n \}\{d \} \), and \( H^d_t = \{ x \} \). If \( D_{it} \ni d, \sigma_{it} = 0 \), then at \( t' > t, j \)'s best response is \( D_{jt'} \ni d, \sigma_{jt'} = 1 \). Thus, at \( t, i \) prefers \( D_{it} \) s.t. \( d \notin D_{it} \).

Hence, in any sequentially rational strategy profile, no arguments are made. ■

The core intuition for this result is the balance of counter-arguments. Under this debate rule, one’s choice of counter-arguments by choosing a speech with the opposite value on a previously argued dimension always comes. If a previous oppositely biased speaker moved the audience toward herself by making an argument on the last of the argument dimensions (recall that the number of dimensions is finite), then one loses nothing and gains in expectation by “counter-arguing” - making the opposite argument on the same dimension. But given Lemma 2, this would make the previous speaker worse off and so deters her from making her argument on that dimension. This logic, however, applies with respect to the next-to-last dimension as well, and so on, to the first dimension. Thus, on the equilibrium path of play, no arguments will be made.

Sequential Fixed-end Speech

Proposition 4 (1) If the proportion of B-agents among the listeners is high enough or all the speakers are sufficiently moderate with respect to the audience, then no arguments are made in equilibrium;

(2) If the proportion of B-agents among the listeners is low enough or some speakers are sufficiently extreme with respect to the audience, then all possible arguments are made in equilibrium;

(3) In the intermediate cases, some arguments are made, but those arguments cover only
a proper subset of argument dimensions and are all biased in the direction of the bias of the last speaker in the speaking order.

**Proof** (sketch) Fix the speaking order $O$.

(1) If $b$ is high enough, then the claim follows immediately by Lemma 2. So, assume that for some level of ideological moderation among the speakers, $b$ is low enough to interest at least some speakers in making arguments if they are speaking alone. Let $i$ be the last speaker who would make speech to $R$ as the only speaker. Suppose that $i$ is followed by a speaker $j$ with a bias in the opposite direction, relative to the audience. Because $j$ will offer the opposite speech of $i$ on every dimension $k$ for which $s_i^k \neq x$, in expectation the audience’s positions will not change. From Lemma 2, $i$ then prefers $s_i = (x, x, ..., x)$, and so no arguments are made.

(2) Suppose the last speaker is a relative extremist who would make arguments on all dimensions or on enough dimensions that some other speaker would prefer that all possible arguments be made than just those that the last speaker would offer. Then some speaker will offer speech on all dimensions with the opposite bias of the final speaker and the final speaker will make speech with her preferred bias on all dimensions.

(3) Suppose the last speaker would make speech to $R$ as the only speaker, and that all other speakers prefer the expected outcome of that speech to the expected outcome of all arguments being made. Then the only speech in equilibrium will be that which the last speaker would choose. To see this, suppose some speaker makes speech with the opposite bias to that preferred by the last speaker, or with the same bias but too extreme. Then, in the cases of the former, the last speaker will offer the opposite speech on the same dimensions, in expectation restoring the initial distribution of policy positions, and the additional speech in the direction of bias that she prefers. In the case of the latter, she makes the opposite arguments on a subset of the dimensions on which the earlier speaker made persuasive speech - a number equal to the excess over what she would have chosen. In both cases, the expected distribution of policy outcomes is the same as if only the last speaker made arguments. Thus, by Lemma 2, the earlier speaker is better off choosing uninformative speech.

The equilibria for this debate rule underscore the fundamental role of the fixed speaking order in affecting the informational content of debate. They suggest that we are likely to see more speech under this rule than under the open-ended sequential speech rule. They also indicate that the order of speakers can have a substantial effect both the informativeness of debate and on the overall ideological posture of the arguments aired in it.

**Discussion**

Our results suggest that the context of public politics provides participants in the debate with stronger incentives to make (informative) arguments. Indeed, the effects of this mode of decision-making are so strong as to dominate the variations in the incentives provided by the different debate rules and by the composition of the audience with respect to both ideology and cognition (as measured by the proportion of Bayesian agents among them). In contrast, the context of private politics creates considerable variation in expected outcomes of deliberation in relation to debate rules and the composition of the audience. We show
that given any positive likelihood of NB-agents in the audience, the simultaneous speech rule may be particularly attractive in bringing about fully informative debate. Perhaps somewhat surprisingly, sequential open-ended debate, which, arguably, captures the deliberative democratic ideal of free communication most closely, may be least desirable from the standpoint of the informativeness of debate.

Appendix

Lemma 1

Proof (1) Let \( l_i \) be the number of “1”s and \( m_i \) be the number of “0”s that \( i \) learns from the speaker.

An immediate implication of the fact that the NB-receivers make no inferences from “null” observations is that they treat these newly “learned” arguments as though they were, effectively, randomly revealed. Thus their posterior beliefs are characterized by

\[
p(\theta_i | l + l', m + m') = \frac{\theta^{l+l'} (1-\theta)^{m+m'}}{\int_0^1 \theta^{l+l'} (1-\hat{\theta})^{m+m'} d\theta}.
\]

(2) Because the B-agent obtains the same information from \( s_i^k = 1 \) and \( s_i^k = 0 \), and listeners and speakers have common prior beliefs about the likelihood that listeners will be persuaded by any given argument, the B-agent’s position is, in expectation, unaffected by \( \sigma \).

Lemma 2

Proof First, we show that the speaker is worse off if the listener learns one dimension than if she learns none. We then show that the speaker is worse off if the listener learns \( d + 1 \) dimensions of \( r \) than if she learns \( d \).

Given that \( \pi^*_i = E[\theta_i | p(\theta | \cdot)] \), listener \( j \) chooses \( \pi^*_j = E[\theta_j | p(\theta | l'_j, m'_j)] \) if the speaker chooses \( (x, x, \ldots, x) \). Suppose listener \( j \) learns one dimension (i.e., \( n = l'_j + m'_j + 1 \)). She chooses \( \pi^*_j = E[\theta_j | p(\theta | l'_j + 1, m'_j)] \) if she infers argument “1,” and \( \pi^*_j = E[\theta_j | p(\theta | l'_j, m'_j + 1)] \) if she infers argument “0.” The probability of her learning “1” is \( \theta_j \), so the expected probability of “1” is \( E[\theta_j | p(\theta | l'_j, m'_j)] \). Her expected probability of “0” is \( E[1 - \theta_j | p(\theta | l'_j, m'_j)] \). Let \( l'_i, m'_i \) be speaker \( i \)’s known numbers of ones and zeroes, respectively. \( i \)’s expected utility from \( j \)’s
choice if she is silent is greater than if she speaks iff

\[ - \int_0^1 p(\theta|l', m_j') \left( E[\theta_j|p(\theta|l'_j, m_j')] - \theta \right)^2 d\theta \]

\[ > -E[\theta_j|p(\theta|l'_j, m_j')] \int_0^1 p(\theta|l'_i, m_i') \left( E[\theta_j|p(\theta|l'_j + 1, m_j')] - \theta \right)^2 d\theta \]

\[ - (1 - E[\theta_j|p(\theta|l'_j, m_j')]) \int_0^1 p(\theta|l'_i, m_i') \left( E[\theta_j|p(\theta|l'_j, m_j' + 1)] - \theta \right)^2 d\theta. \]

Taking expectations, we obtain an equivalent inequality

\[ - (E[\theta_j|p(\theta|l_j', m_j')])^2 + 2E[\theta_j|p(\theta|l_j', m_j')] E[\theta_i] - E[\theta_i^2] \]

\[ > -E[\theta_j|p(\theta|l_j', m_j')] \left( (E[\theta_j|p(\theta|l_j' + 1, m_j')])^2 - 2E[\theta_j|p(\theta|l_j' + 1, m_j')] E[\theta_i] + E[\theta_i^2] \right) \]

\[ - (1 - E[\theta_j|p(\theta|l_j', m_j')]) \left( (E[\theta_j|p(\theta|l_j', m_j' + 1)])^2 - 2E[\theta_j|p(\theta|l_j', m_j' + 1)] E[\theta_i] + E[\theta_i^2] \right). \]

Because

\[ E[\theta_j|p(\theta|l_j' + 1, m_j')] - E[\theta_j|p(\theta|l_j', m_j')] < E[\theta_j|p(\theta|l_j', m_j')] - E[\theta_j|p(\theta|l_j', m_j' + 1)] \]

iff \( E[\theta_j|p(\theta|l_j', m_j')] > \frac{1}{2} \), (3) is always true.

Suppose now that the speaker prefers the listener learning nothing to learning \( k \) dimensions. Define \( l'' \) and \( m'' \) such that \( l'' \geq l', m'' \geq m' \), and \( (l'' - l') + (m'' - m') = k \). Substituting \( l'' \) for \( l' \) and \( m'' \) for \( m' \) in equation (3) yields the condition for the speaker preferring the listener’s learning only \( k \) dimensions to her learning \( (k + 1) \). Because this argument is independent of the group composition and the speaker’s type, it must hold for any group of Bayesian listeners.

Because only two arguments are possible, \( \{0, 1\} \), a B-agent learns her argument on dimension \( k \) from any \( s^k \in \{0, 1\} \). In contrast, an NB-agent is certain to learn her argument on dimension \( k \) iff she hears both arguments on that dimension. ■

References


