# Deliberation and Voting Rules[1]

David Austen-Smith
Departments of Political Science and Economics
Northwestern University
Evanston, IL 60208

Tim Feddersen
MEDS, Kellogg Graduate School of Management
Northwestern University
Evanston, IL 60208

Revised, October 16 2002

[1]Paper prepared for the Wallis Institute of Political Economy annual conference on political economy, University of Rochester, NY, 10/25/02-10/26/02. We have benefited greatly from deliberating on deliberation with Randy Calvert, Daniel Diermeier and Michael Wallerstein. The deficiencies in the paper, however, are ours alone.

**Abstract**

We analyse a formal model of decision-making by a deliberative committee. There is a given binary agenda. Individuals evaluate the two alternatives on both private and common interest grounds. Each individual has two sorts of private information going into committee: (a) perfect information about their personal bias and (b) noisy information about which alternative is best with respect to a (commonly held) normative criterion. Prior to a committee vote to choose an alternative, committee members engage in deliberation, modeled as a simultaneous cheap-talk game. We explore and compare equilibrium properties under majority and unanimity voting rules, paying particular attention to the character of debate (who influences who and how) and quality of the decision in each instance. On balance, majority rule induces more information sharing and fewer decision-making errors than unanimity. Furthermore, the influence and character of deliberation *per se* can vary more under majority rule than under unanimity.

# 1 Introduction

The importance of deliberation for social choice has long been recognized. One example, here as in so many areas of voting theory, is Condorcet. Condorcet saw the role of deliberation and debate largely in positive terms, as necessary both to clarify individual interests and to formulate coherent agendas over which to vote: he writes

> "Discussions in a debating assembly clearly have two main concerns. First, there is a discussion about principles fundamental to any decision on a general question ... This is followed by another debate [in which the general question] can be reduced to a number of clear and simple questions about which the assembly can be consulted. If this reduction is done perfectly, then each individual can give a true expression of his will by replying *yes* or *no* to each of these basic questions. ... The first [kind of discussion] is sufficient for men who simply want to clarify their ideas and form an opinion, while the second is of use only to men who are required to prepare or pronounce a joint decision. ... Without prior discussion in an assembly established for this purpose, it would be virtually impossible to prepare motions, or to present them in such a way as to permit an immediate decision either by this assembly or by any other." (Marquis de Condorcet, 1793, as translated by Iain McLean and Fiona Hewitt, 1994:193)

Although much of the recent literature on so-called "deliberative democracy" is more expressly normative, being concerned with questions of legitimacy and achieving a consensus sufficient to make voting irrelevant,[1] there is occasionally some recognition that reality is likely to fall short of the ideal:[2]

> "[I]deal deliberation aims to arrive at a rationally motivated consensus — to find reasons that are persuasive to all who are committed to acting on the results of a free and reasoned assessment of alternatives by equals. Even under ideal conditions there is no promise that consensual reasons will be forthcoming. If they are not, then deliberation concludes with voting, subject to some form of majority rule. The fact that it may so conclude does not, however, eliminate the distinction between deliberative forms of collective choice and forms that aggregate non-deliberative preferences. The institutional consequences are likely to be different in the two cases, and the results of voting among those who are committed to finding reasons that are persuasive to all are likely to differ

[1] For example, see the contributions to Bohman and Rehg, 1997, and to Elster, 2000.

[2] Not all normative theorists writing on deliberative democracy are so enthusiastic about the value of the process. Particularly coherent critiques are offered by Christiano (1997), Johnson and Knight (1997) and Sanders (1997).

from the results of an aggregation that proceeds in the absence of this committment." (Joshua Cohen, 1989:23)

The quotations above reflect different concerns but a common interest in deliberative committees. It seems clear that both Condorcet and Cohen, along with almost all of those writing on deliberative democracy, see information sharing and conceptual discussion as being necessary for effective or legitimate collective decision-making. Exactly what such information sharing or conceptual discussion, that is, what deliberation entails or implies for social choice is less clear. How might the incentives for deliberation depend on details of the rules governing deliberation or on details of the voting procedure used to reach a final decision? Are decisions made following deliberation always at least weakly better than those made in its absence, relative to some widely accepted normative criterion? Does it, or should it, matter what sort of arguments are deemed admissible in collective deliberation? Does all deliberation involve information sharing or can there be productive and influential deliberation that is not fundamentally informational? Although we do not pretend to answer any of these questions definitively here, we nevetheless hope at least to understand better how voting rules influence information sharing and, in so doing, offer some insight on the role of different sorts of reason in debate.

Throughout, our focus is on *deliberative committee decision-making*, where committee membership is at least two and completely describes both the set of individuals involved in any deliberation over a collective decision and the set of individuals responsible for making such a decision. It is useful to distinguish deliberative committee decision-making from *hearings*, that is, from settings in which some list of relatively informed agents give advice or offer testimony to a relatively uninformed agent who unilaterally makes a decision (e.g. Diermeier and Feddersen, 2000). To the extent that deliberation involves strategic information transmission – and much, if not all, of it certainly does – the literature concerned with hearings is clearly germane.[3] For instance, Glazer and Rubinstein (2001) show that a given (and truthful) argument can function quite differently if offered in support of the decision-maker choosing a particular decision rather than as a counterargument to some opposing advisor's claims and, moreover, that such a feature is characteristic of any debate in their setting that minimizes the likelihood of an ex post error from the decision-maker's perspective; Matthews and Postlewaite (1995) and Austen-Smith (1993a, 1993b) provide examples to illustrate that order of speaking in a multiple sender cheap talk game can matter a great deal; Lipman and Seppi (1995) study debates between fully informed senders capable of offering partially provable arguments to the decision-maker; Ottaviani and Sorensen (2001) consider herding problems induced by a sequence of asymmetrically informed advisors with

[3]Indeed, almost all of the strategic information transmission papers with "deliberation" or "debate" in the title fall within this category; see Glazer and Rubinstein (2001), Ottaviani and Sorensen (2001), Spector (2000).

reputational concerns; and Spector (2000) analyses a quite unusual dynamic model in an effort to understand why, in a multi-dimensional world, so much political conflict appears one-dimensional. Although some of what we have to say exploits the literature, we are not here especially concerned with hearings but with deliberative committees. Deliberative committees are of widespread empirical relevance; they raise questions of consensus, conversation and institutional design of marginal relevance to hearings; and, from a strategic perspective, little is yet known about the implications or properties of deliberation for collective choice.[4]

Before going on to consider some of the issues in a specific context, we briefly identify two important themes in the normative literature on deliberation, both of which are flagged in the quotation above from Cohen. The first theme, consensus, is more often than not seen as a goal or ideal for any deliberative process while the second, the legitimacy of reasons, is an integral part of the process itself.

## 2 Consensus and reasons

Although what is entailed by a deliberation generating or seeking a consensus is rarely made precise, at least three sorts of logically distinct domains or meanings of consensus can be found in the political theory literature: preference, informational and justificatory.

*Preference consensus.* The strongest of the three notions is consensus in preferences. The claim is that deliberation is a "transformative" process, a process that changes in individuals' primitive preferences over outcomes sufficiently to yield complete agreement on the collective decision. So, while Pareto efficiency is implied by preference consensus, efficiency *per se* is not enough. That individuals' primitive preferences might in fact change and evolve over time is likely the case. But exactly how such transformations are thought to occur through deliberation over collective choice, or how they might be distinguished empirically from, in particular, changes in beliefs about the consequences of choices, is at best obscure and we have no more to say about the issue here. On the other hand, many of the references to "transforming" or "changing" preferences through deliberation (e.g. Manin, 1987; Cohen, 1989; Miller, 1993; Sunstein, 1993) can be readily understood when phrased in terms of *induced* rather than primitive preferences. From this perspective, it is induced preferences over (collective) actions that are subject to change, not primitive preferences over the consequences of such actions, and in principle it seems quite possible for communication and conversation to result in at least some degree of consensus in induced preferences defined over available actions. Such a focus points to more familiar analytical territory and leads us to the second conception of deliberation.

[4] There is currently very little in the strategic literature on deliberative committees. Exceptions include Austen-Smith (1990a, 1990b), Calvert and Johnson (1998), Coughlan (2000) and Doraszelski, Gerardi and Squintani (2001).

*Informational consensus.* On the informational account, any "consensus" derived from deliberation is at best consensus on the likely consequences of taking any action; because primitive preferences are assumed fixed, disagreement about the "right" action to take can persist when there are common beliefs about the consequences of any action.[5] Such a notion of consensus appears to be an essential part of all accounts of deliberative democracy and has been central to almost all of the formal literature concerned with communication and collective decision-making. Whether it is the case that such consensus will in fact materialize through deliberation, however, is not clear and, unlike the primitive preference concept,we have something to say on this.

*Justificatory consensus.* Finally, justificatory consensus is intimately related to the legitimacy of reasons. Here, participants may come to agree on a collective decision, not so much because they happen to like it or personally believe it is the best, but because they are unable to present a sufficiently cogent and publicly legitimate argument for any alternative decision. For example, juries must convict or acquit on the basis of legally permissible evidence. Thus, jurors can simultaneously disagree about whether a defendant actually committed the crime yet all agree that he is legally innocent. It is possible, then, for deliberation to result in justificatory consensus without there being any preference or informational consensus: an individual may fundamentally disagree with the proposed decision on grounds at least partly predicated on private information, yet recognize that any argument revealing this information would be deemed either insufficient or illegitimate with respect to the commonly held norms of debate. And this leads to the second concern of the section, the role of reasons in deliberation.

What is required for an argument or a reason to be 'publicly legitimate' is the subject of much of the normative literature (e.g. Gutman and Thompson, 1995; Cohen 1989; Estlund, 1997; Gauss, 1997) and, save in regard to issues of credibility and equilibrium refinement through beliefs, has attracted no explicit interest among social choice or game theorists. It is unnecessary, at least for current purposes, to tackle the question of legitimacy here. Instead, it suffices to note that a central (perhaps the central) characteristic of a legitimate reason for collective is decision that it is a reason grounded on some concept of the "common good"; in particular, self-regarding reasons are deemed illegitimate in public deliberation. Thus legitimate positive arguments on the consequences of making a particular decision might be purely informational in the sense captured in the standard models of incomplete information and uncertainty, or may depend on analogic and inductive reasoning as sketched out in a recent paper of Aragones et al (2001). Similarly, legitimate normative reasoning might be reasoning on the logical coherence of various principles of justice[6], or on appeals to some

[5] If strategic uncertainty is included here, then deliberation aimed at informational consensus is successful to the extent that committee members coordinate on some joint collective action (Calvert and Johnson, 1998).

[6] In which case, the classical social choice theorems – especially those concerned with interpersonal welfare comparisons – may be interpreted as theorems about what sorts of reason are mutually consistent.

notion of intrinsic value or merit, or on coordinating on one of several possible collective outcomes. In any event, the focus is on argument designed to determine and articulate the alternatives that are (somehow) "best for society". Such arguments, to be legitimate, are held to rest on generalizable principles or values providing, under conditions of "ideal deliberation ... reasons that are persuasive to all who are committed to acting on a free and reasoned assessment of alternatives by equals" (Cohen, 1989). On the other side, arguments designed to form a winning voting coalition on the grounds that the particular coalition maximizes that group's aggregate or individual payoffs at the expense of some minority are not, on this account, legitimate. Likewise, arguments predicated at least in part on treating individuals unequally on grounds of race, religion or any other morally irrelevant criterion are deemed illegitimate.

The extent to which deliberative reasoning can be purged of self-interested motivation is, however, unclear, even under conditions of "ideal deliberation" (Elster, 1997). On the one hand, the literature on strategic information sharing suggests that arguments which, if believed, result exclusively in the speaker's self-interest being furthered, typically carry no weight, rendering moot the normative concern to exclude them as illegitimate; and on the other hand, however sincerely a speaker might offer a "common good" argument, his or her audience cannot be compelled to interpret the argument without some inference on the speaker's private interests. Thus, being obliged to suppress any explicit statement of personal self-interest in a collective choice does not imply that any "common good" arguments are in fact uncoloured by such self-interest.

With the preceding remarks in mind, the rest of the paper is devoted to considering deliberation in a formal model of committee decision making. Although, for Condorcet at least, the most important role of deliberation is perhaps in agenda-setting, we assume there is an exogenously fixed agenda.[7] Although it is fairly natural to begin by asking what happens with fixed alternatives and then back up to ask how the alternatives for consideration might be chosen, the Condorcet Jury Theorem along with recent results on information aggregation through voting over fixed binary agendas (e.g. Ladha, 1992; McLennan, 1998; Duggan and Martinelli, 2001; Feddersen and Pesendorfer, 1996, 1997) raise a more concrete question about whether deliberation over given agendas is a salient issue. At least asymptotically as the electorate gets large, any majority or supermajority voting rule short of unanimity almost surely selects the alternative that would be chosen under the given rule were all voters fully informed and surely voted. However, committees in which deliberation is feasible are typically too small for asymptotic results to be useful.[8]

[7] The only explicit model of deliberative committee decision making with an endogenous agenda of which we know is Austen-Smith (1990a, 1990b). Comparing the results there with those developed here, it seems the role and character of debate with endogenous agenda-setting can exhibit some very different characteristics to those with a fixed agenda.

[8] Condorcet surely felt this to be significant when claiming that the second form of debate he identifies (in which general questions are refined into "a number of clear and simple questions") "could not take place

Thus there remains room for decision-relevant information sharing and argument in committees and we therefore consider two of the questions about committee deliberation raised earlier: the implications of the decision rule on the character of information sharing; and the nature of reasons offered in deliberation. In turn, addressing these issues yields insight on (some of) the various notions consensus identified earlier.

## 3 A deliberative committee

Consider a three person committee, $N = \{1, 2, 3\}$, that has to choose an alternative $z \in \{x, y\}$. Individual preferences over the feasible alternatives can be decomposed into two parts, one reflecting purely private interests and one reflecting a notion of common good or fairness. Specifically, for any $i \in N$, $i$'s private interests are given by a utility

$$u_i(x) = 1 - u_i(y) \in \{0, 1\};$$

let $b_i \in \{x, y\}$ be $i$'s *bias*, where $b_i = z$ if and only $u_i(z) = 1$. The common good value of an alternative $z \in \{x, y\}$ is $f(z|\omega) \in \{0, 1\}$, describing which alternative is fair in state $\omega \in \{X, Y\}$. Then for any $z \in \{x, y\}$, $b_i \in \{x, y\}$ and $\lambda \in [0, 1]$, assume $i$'s preferences can be represented by[9]

$$U(z; b_i) = \lambda u_i(z) + (1 - \lambda) f(z|\omega).$$

In general, different individuals can be expected to have different moral systems or senses of what constitutes the common good. For example, suppose individuals are either Benthamite Utilitarians or Rawlsian Maximinimizers. Then reasons for choosing one alternative over another that are germane to the former can be utterly irrelevant to the latter and conversely. In this setting, productive debate might proceed either by a discussion of principles along, say, axiomatic grounds, or by seeking out reasons and arguments that are decision-relevant to both conceptions of how to evaluate the common good. Although such issues are, we think, quite important and worth thinking about more deeply, for now it is convenient simply to ignore such differences. So assume the evaluation function $f$ is the same for everyone and satisfies $f(z|\omega) = 1$ if and only if $\omega = Z$. Similarly, without suggesting the assumption describes reality, it is convenient to suppose individuals value the common good in the same way, so $\lambda$ is common across committee members.

There are two substantive sources of incomplete information. First, individual $i$'s bias $b_i \in \{x, y\}$ is known only to $i$: for all $i \in N$, assume the probability that $b_i = x$ is $1/2$. The second informational incompleteness concerns which of the two alternatives is most in the common interest, modeled as uncertainty over the realized state $\omega \in \{X, Y\}$. The

outside an assembly without becoming very time-consuming" and "is of use only to men who are required to prepare or pronounce a joint decision" (*ibid*).

[9] See Karni and Safra (2002) for an axiomatic justification of separable preferences for individuals with both private interests and a preference for fairness.

common prior belief over $\{X, Y\}$ is assumed uniform. With probability $(1-q) \in (0,1)$ an individual $i \in N$ is either *uninformed*, observing no further information denoted $s_i = 0$, or, with probability $q$ is *informed* and observes a noisy signal $s_i \in \{-1, 1\}$ from a common state-dependent distribution. Whether or not any $i \in N$ has observed any signal and, if so, which signal he or she received, is private information to $i$. Conditional on observing a signal, let

$$p = \Pr[s_i = 1 | X, s_i \neq 0] = \Pr[s_i = -1 | Y, s_i \neq 0]$$

and assume $p \in (\frac{1}{2}, 1)$. A pair $(p, q) \in (\frac{1}{2}, 1) \times (0, 1)$ is called an *information structure.*

In sum, therefore, an individual's type going into the committee decision-making process is a pair $(b, s)$ where $b \in \{x, y\}$ is the alternative most in the individual's private interests and $s \in \{-1, 0, 1\}$ is the individual's signal regarding which alternative is fair. Hence, for each alternative $z \in \{x, y\}$, $i \in N$ has induced preferences going into committee given by

$$E[U(z; b_i)|s] = \lambda u_i(z) + (1-\lambda)\Pr[Z|s] \in [0, 1].$$

Clearly, if $\lambda > 1/2$ then no type ever cares sufficiently about the common good for it to be decision relevant; therefore assume hereon that $\lambda \in (0, 1/2)$. Let $\pi \equiv \Pr[\omega = X]$ and define

$$\begin{aligned} \pi_x(\lambda) &= \min \{\pi : \lambda + (1-\lambda)\pi \geq (1-\lambda)(1-\pi)\}; \\ \pi_y(\lambda) &= \min \{\pi : (1-\lambda)\pi \geq \lambda + (1-\lambda)(1-\pi)\}. \end{aligned}$$

Then $\pi_x(\lambda)$ [respectively, $\pi_y(\lambda)$] is the decreasing [respectively, increasing] curve in Figure 1 below, illustrating induced preferences in $(\pi, \lambda)$-space. If $\pi < \pi_x(\lambda)$, an $x$-biased individual (i.e. $i$ such that $b_i = x$) nevertheless strictly prefers $y$ to $x$ on grounds of expected fairness and, similarly, if $\pi > \pi_y(\lambda)$ then a $y$-biased individual (i.e. $i$ such that $b_i = y$) strictly prefers $x$ to $y$. The more individuals focus on their private interests (the higher is $\lambda$), the more evidence on the relative fairness of the two alternatives they require for such interests to be dominated.

It is analytically useful to define critical values for $\lambda$, $l_1(p), l_2(p) \in (0, 1/2)$, by

$$l_1(p) \equiv \frac{2p-1}{2p} \textit{ and } l_2(p) \equiv \frac{1}{p} l_1(p)$$

where there is no ambiguity, write $l_1 = l_1(p)$, etc. To interpret $l_1(p)$, let $S \equiv \sum_{i \in N} s_i$ be the sum of all individuals' signals. If this sum were common knowledge and if $\lambda < l_1$, then all individuals strictly prefer $x$ [respectively, $y$] when $S \geq 1$ [respectively, $S \leq -1$]. And for $\lambda \in (l_1, l_2)$, all individuals strictly prefer $x$ [respectively, $y$] only when $S \geq 2$ [respectively, $S \leq -2$]. When $\lambda$ is below the threshold $l_1$, individuals' induced preferences and behaviour in committee are in principle most sensitive to the opportunities offered by deliberation. From this perspective increases in $p$ at a given $\lambda$ are analogous to reductions in $\lambda$ at a given $p$. Hence it suffices for the most part to focus on $\lambda < l_1$.

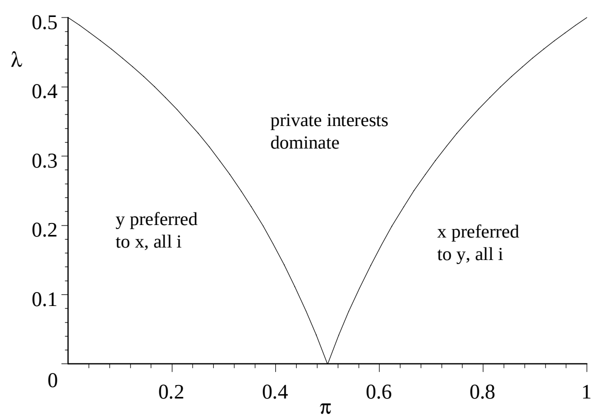


Figure 1: Induced preferences

Once types are fixed, the committee decision-making process has two stages: the final stage is a vote (with no abstention) to choose between the two alternatives; this may be preceeded by a "debate" in which committee members simultaneously send a cheap talk message about the committee choice.

### 3.1 Strategies and equilibria

The solution concept is a refinement of Perfect Bayesian Equilibrium in undominated (anonymous) strategies; although details of the refinement are discussed later, any subsequent reference to "equilibrium" or "equilibrium behaviour" refers to this solution concept. Anonymous strategies are imposed by definition and assumed throughout; in effect, anonymous strategies do not depend on the names of the agents.

A message strategy is a map

$$\mu : \{x, y\} \times \{-1, 0, 1\} \to \mathcal{M}$$

where $\mathcal{M}$ is an arbitrary uncountable list of messages or speeches. Thus, any individual $i \in N$ with bias $b_i$ and signal $s_i$ makes a speech (sends message) $\mu(b_i, s_i) \in \mathcal{M}$. Let

$$\mathcal{M}_\mu \equiv \left[\cup_{\{x,y\}\times\{-1,0,1\}} \mu(b, s)\right] \subseteq \mathcal{M}$$

denote the range of $\mu$. A *debate* is a list of messages $\mathbf{m} = (m_i, m_j, m_k) \in \mathcal{M}_\mu^3$; for any $i \in N$ and debate $\mathbf{m}$, let $M_{-i} = \sum_{j \neq i} m_j$.

A (pure) voting strategy is a map

$$\upsilon : \{x, y\} \times \{-1, 0, 1\} \times \mathcal{M}^3 \to \{x, y\}$$

So for any individual with bias $b_i$ and signal $s_i$, $\upsilon(b_i, s_i, \mathbf{m}) \in \{x, y\}$ describes the individual's vote conditional on $b_i$, $s_i$ and the debate $\mathbf{m} = (m_i, m_j, m_k) \in \mathcal{M}^3$; by convention, the message listed first is invariably that of the individual $i$. Let $\mathbf{v} = (v_i, v_j, v_k) \in \{x, y\}^3$ denote a list of votes.

There are two sorts of constraint that any equilibrium strategy pair $(\mu, \upsilon)$ must satisfy. In any equilibrium, rational individuals vote for that alternative they most prefer conditional on their type, on the equilibrium messages heard in debate and on the event that they are pivotal in the vote under the committee decision rule in effect. That is, for any $i \in N$ and signal $s_i$, given the voting behaviour $\upsilon_{-i}$ of committee members other than $i$, given the voting rule and given a message strategy $\mu$ yielding debate $\mathbf{m} \in \mathcal{M}^3$, $\upsilon$ must satisfy:

$$E[U(z; b_i)|s_i, \mathbf{m}, \mu, z, \upsilon_{-i}, \mathbf{votepiv}] > E[U(z'; b_i)|s_i, \mathbf{m}, \mu, z', \upsilon_{-i}, \mathbf{votepiv}]$$

implies $i$ surely votes $v_i = z$ rather than $v_i' = z'$, where **votepiv** denotes the event that $i$'s vote is pivotal at the voting stage. Consequently,the first set of constraints, the *pivotal voting constraints*, insure that all individuals' voting behaviour is optimal conditional on their vote being pivotal at every information set. In some cases, such voting recommends voting with an individual's private interests and it is useful to have a term for this: say an individual *votes her bias* if $b_i = z$ implies $i$ surely votes $z$. On the other hand, say that $i$ *votes her signal* if, irrespective of bias, $s_i = 1$ [respectively, $s_i = -1$] implies $i$ votes for $x$ [respectively, $y$].

The second set of constraints, the *pivotal signaling constraints*, insure that every individuals' message is optimal conditional on that message being pivotal for the final committee outcome, given individuals voting strategies. Specifically, for any $i \in N$ and signal $s_i$, given the voting strategy $\upsilon$, the voting rule, and message strategies $\mu_{-i}$ for individuals other than $i$, $\mu$ must satisfy:

$$E[U(z; b_i)|s_i, m_i, \mu_{-i}, \upsilon, \mathbf{sigpiv}] > E[U(z'; b_i)|s_i, m_i', \mu_{-i}, \upsilon, \mathbf{sigpiv}]$$

implies $i$ surely makes the speech $m_i$ rather than the speech $m_i'$, where **sigpiv** denotes the event that $i$ is pivotal at the signaling (debate) stage.

Satisfying both sets of constraints gives rise to a variety of equilibria. We focus on equilibria involving three important forms of debate. A (pure) message strategy $\mu$ is:

*Separating in common interest* if, for all $b \in \{x, y\}$ and any distinct $s, s' \in \{-1, 0, 1\}$,

$$\mu(b, s) \neq \mu(b, s');$$

*Semi-pooling in common interest* if, for all $b \in \{x, y\}$,

$$\mu(x, 0) = \mu(b, 1) \neq \mu(b, -1) = \mu(y, 0);$$

and *Pooling in common interest* if, for all $s, s' \in \{-1, 0, 1\}$,

$$\mu(x,s) = \mu(x,s') \text{ and } \mu(y,s) = \mu(y,s').$$

Further, say that $\mu$ is *separating in private interests* if, for all $s \in \{-1,0,1\}$, $\mu(x,s) \neq \mu(y,s)$.

An equilibrium $(\mu, \upsilon)$ is said to be a *separating debate equilibrium* if $\mu$ is separating in common interest; analogously, define semi-pooling and pooling debate equilibria.

Because debate is cheap talk there is always an equilibrium in which no information is revealed in debate (Farrell, 1993). Further, private interest information (bias) matters in debate only insofar as it influences the audience's interpretation of any information offered regarding the common good. Hence, there is always an equilibrium in which message strategies are separating in private interests but pooling in common interests. To see this, let $(\mu, \upsilon)$ be any equilibrium in which $\mu$ is pooling in common interests; partition the message space $\mathcal{M}$ into two sets, $\mathcal{M}_x$ and $\mathcal{M}_y$, and define the message strategy $\mu'$ such that, for all signals $s$, $\mu(x,s) \in \mathcal{M}_x$ and $\mu(y,s) \in \mathcal{M}_y$. Then the debate $\mathbf{m} \in \mathcal{M}^3_{\mu'}$ fully reveals the bias distribution but contains no more information on common interests than does the pooling strategy $\mu$. Therefore, since $(\mu, \upsilon)$ is an equilibrium, the strategy pair $(\mu', \upsilon)$ is likewise an equilibrium.

A more interesting question concerns the extent to which more informative debates with respect to common interest might also involve information with respect to bias. In this regard, if there exists an equilibrium with $\mu$ separating in common interests, then speakers are able to include a credible statement of their private interests too: information on bias matters only insofar as it influences the interpretation of speeches on the common good and, under separation in common interest, all of the decision-relevant information is shared. On the other hand, the same is not true if message strategies are semi-pooling in common interests; here, a listener's equilibrium interpretation of a message depends essentially on his or her beliefs regarding the speaker's private bias so statements regarding individual bias become consequential. This suggests a speaker might try to influence any interpretation of her speech by offering information about her private interests along with information about the common good. But it is not hard to see that such elaboration cannot be persuasive: if an informed speaker can convince others of the truth of her claims about the common good by saying something about her bias, an uninformed individual could make the same speech and also be convincing.

Messages in the model are cheap talk and may therefore have any form at all. Nevertheless, it is convenient to think about them as natural language arguments about what to choose. The preceding remarks imply that it suffices (at least for the formal analysis) to consider only the common interest content of any speech or debate: under separating or pooling message strategies, discussion of private interests is irrelevant and under semi-pooling message strategies such discussion is relevant but impossible. Thus speeches have essentially three possible decision-relevant interpretations and there is no loss of generality in taking $\mathcal{M} =$

$\{-1, 0, 1\}$ hereafter. The literal content of the messages $m = -1$, $m = 0$ and $m = 1$ are, respectively, the speeches "I believe $y$ is likely the best choice", "I am uninformed" and "I believe $x$ is likely the best choice".

We wish to understand how deliberation influences subsequent voting behaviour and thereby the quality of committee decisions. It is not enough simply to look for equilibria exhibiting more or less informative signaling strategies: any given informative signaling strategy can in principle be consistent in equilibrium with many voting strategies. Although some of the variation in voting behaviour at any given parameterization is eliminated through refinement, some remains and is substantively interesting. Perhaps not surprisingly, given any attitude toward fairness ($\lambda$), the influence of deliberative argument on voting depends on the distribution of private bias in the committee and the relative likelihoods of any individual being informed ($q$) and the quality of the information conditional on being informed ($p$). Exactly how these features of the environment interact and the character of deliberative influence that they support, however, is not immediately apparent and turns out to be quite subtle. More detailed discussion of the particular sorts of influential equilibrium behaviour that can arise is deferred until the analysis, and we conclude this section with a brief description of the refinement (the formal definition is given in Appendix B).

The possibility, at any given parameterization of the model, of out-of-equilibrium messages and of undominated equilibrium voting profiles under which no individual is pivotal, motivates using a refinement to sharpen predictions. The refinement is essentially technical and has two components. First, individual vote decisions are subjected to individual-invariant trembles and we report behaviour in the limit as the trembles become vanishingly small. This insures that all individuals' equilibrium strategies are the limit of a sequence of best response strategies chosen conditional on being pivotal with strictly positive probability. The second component is a restriction on out-of-equilibrium beliefs at the signaling stage. The issue here arises only for semi-pooling debate equilibria in which uninformed individuals are supposed to speak in suppport of their bias. There is little guidance on how best to proceed here and we simply assume that listeners hearing an out-of-equilibrium speech "I am uninformed" treat it as equivalent to hearing the speech "I believe $y$ is likely the best choice" (where choosing $y$ is without loss of generality).[10]

Although the comparative equilibrium properties of deliberative committee decision making with majority and unanimity voting are a main concern, it is necessary (and of independent interest) to analyse behaviour under the two rules separately. Begin with majority rule.

[10] We also considered an alternative specification: that out-of-equilibrium speeches claiming no information were believed surely. Although there turn out to be some differences, they are inconsequential; the derivations required, however, are considerably more tedious!

# 4 Majority rule

Under majority rule, the alternative receiving at least two votes at the voting stage is the committee decision. This rule is inherently symmetric and, therefore, we consider symmetric strategies here. For any bias $b \in \{x, y\}$ let $-b \equiv \{x, y\} \backslash \{b\}$; recall $s \in \{-1, 0, 1\}$ and $(m_i, m_j, m_k) \in \{-1, 0, 1\}^3$.

**Definition 1** *(1) A message strategy $\mu$ is symmetric if and only if, for all $(b, s)$, $\mu(b, s) = -\mu(-b, -s)$.*

*(2) A vote strategy $\upsilon$ is symmetric if and only if, for all $(b, s, m, m', m'')$,*

$$\upsilon(b, s, m, m', m'') = x \Leftrightarrow \upsilon(-b, -s, -m, -m', -m'') = y.$$

Imposing symmetry on $\mu$ clearly adds little: by definition, if $\mu$ is separating, semi-pooling or pooling in common interest then $\mu$ is symmetric. Requiring symmetric voting strategies, however, although a mild restriction for the present model, has more bite.

Suppose there is no debate or that the signaling strategy is pooling in common interest (which, given debate is cheap talk here, always constitutes an equilibrium strategy). Then it is easy to check the symmetry assumptions imply that, for all $\lambda < 1/2$, the unique equilibrium voting profile is for all individuals to vote for their most preferred alternative conditional on their private signal and on being pivotal for the decision.

**Proposition 1** *Suppose there is no debate or that $\mu$ is pooling in common interests. Then, up to behaviour on the boundary $\lambda = l_1$, there is a unique symmetric voting equilibrium under majority rule; further, for any $z \in \{x, y\}$, if:*

*(1) $\lambda > l_1$ then, for all $s \in \{-1, 0, 1\}$, $\upsilon(z, s, \emptyset) = z$;*

*(2) $\lambda < l_1$ then $\upsilon(z, 1, \emptyset) = x$, $\upsilon(z, -1, \emptyset) = y$ and $\upsilon(z, 0, \emptyset) = z$.*

A proof for Proposition 1 and all subsequent results (save Proposition 5, where the proof is by example) are collected in Appendix A.

Say that an alternative $z$ is the "right" decision (relative to majority rule) if $z$ is an alternative that is preferred by a majority of individuals conditional on fully shared information. An alternative definition of the "right" decision is the alternative most likely in the common interest, conditional on the realized list of signals. When $\lambda < l_1$ the two definitions recommend the same alternative but in general they are distinct. Unlike the definition in terms of full information majority preference, the definition in terms of common interest alone is insensitive to private bias. At least for now, we adopt the majority preference definition. Then, for $\lambda < l_1$, the only event in which the committee decision is not right, is when there are two uninformed individuals with identical bias for $z$ and an informed agent with a signal supporting $z' \neq z$. In this case, all individuals vote for $z'$ under full information but, in equilibrium, a majority votes under private information for $z$. Doing the calculation, the probability of "error" in the committee decision for $\lambda < l_1$ is no bigger than $3q(1-q)^2/8 \leq 1/18$. When $\lambda > l_1$, however, the likelihood of error jumps to 1/2. Now suppose individuals have an opportunity for debate prior to voting.

Because information on common interests is intrinsically imperfect, we abuse language somewhat and say there is "full information" at the voting stage if the realized list of signals $\mathbf{s} \in \{-1, 0, 1\}^3$ is common knowledge. An equilibrium $(\mu, v)$ is a *full information equivalent* equilibrium whenever it surely results in the committee making the right decision. It is worth noting that full information equivalence does not imply all information is revealed in debate but only that, along the equilibrium path, committee decisions are those that would be made under common knowledge that $\mathbf{s} = \mathbf{m}$.

## 4.1 Separating debate equilibria

It is evidently possible for there to exist separating debate equilibria in which deliberation has no impact at all on individual voting behaviour. For example, suppose $\lambda$ is sufficiently high relative to the quality of private information $p$; then no feasible private signal or deliberative argument can outweigh any individual bias and, therefore, fully revealing private information in debate can be an equilibrium strategy precisely because it is inconsequential. More interesting are those separating debate equilibria in which voting behaviour is responsive to deliberation.

In separating debate equilibria, speeches regarding the relative merits of the two alternatives are completely untainted by private interest and deliberation can generate informational consensus, justificatory consensus and, save in the case that common knowledge of signals $\mathbf{s}$ results in all individuals' induced preferences over $\{x, y\}$ being described by their bias, consensus also in induced preferences. While the first and last claims are obvious, the claim that separating debate strategies also yield justificatory consensus may not be so. To see this, recall that under justificatory consensus, no individual has a 'legitimate' argument available to alter a tentative committee decision but need not in fact consider the decision the best one. When all individuals separate in debate, all of the available information is shared and, therefore, there are no available arguments for changing committee members' beliefs and thereby influencing the expected final decision.[11] Finally, and perhaps most importantly, is that if a separating debate equilibrium exhibits full information equivalence, the probability that the equilibrium committee decision is not the right one is zero. In sum, the procedural and consequential properties of full informational equivalent separating debate equilibria are those suggested in the normative literature as central to any conception of legitimate and effective deliberation. Unfortunately, existence of such equilibria cannot always be assured.

[11] If the agenda is not fixed but chosen endogenously by committee members following a debate, then the connection between separating debate strategies and justificatory consensus becomes more obscure. The connection turns out to depend critically on how individuals form beliefs consequent on hearing an equilibrium debate and then observing an out-of-equilibrium proposal added to the agenda. See Austen-Smith 1990b.

**Proposition 2** *Fix an information structure* $(p, q)$*. There is a unique value* $\lambda(p, q) < l_1(p)$ *such that there exists a full information equivalent symmetric separating debate equilibrium if and only if* $\lambda \leq \lambda(p, q)$*. Moreover,*

*(1) for all* $p \in (1/2, 1)$*,* $\lambda(p, q)$ *is strictly single-peaked in* $q$ *on* $(0, 1)$ *with peak* $\lambda^*(p) < l_1(p)$ *such that* $\lim_{p \to 1} \lambda^*(p) = \lim_{p \to 1} l_1(p) = 1/2$*;*

*(2) for all* $q \in (0, 1)$*,* $\lambda(p, q)$ *is strictly increasing in* $p$ *on* $(\frac{1}{2}, 1)$ *with maximum* $\lambda^*(q) < 1/2$ *such that* $\lim_{q \to 1} \lambda^*(q) = 0$*.*

At first glance, statements (1) and (2) of the proposition, taken together, may appear contradictory. However, they simply indicate that the order of limits is consequential: Figure 2 illustrates the function $\lambda(p, q)$ for three values of $q$.

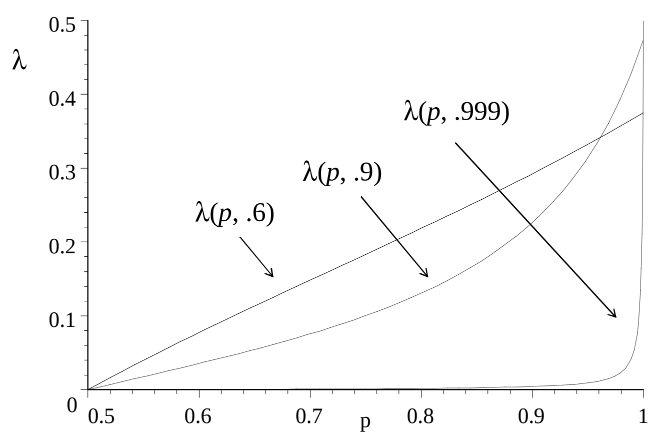


Figure 2: $\lambda(p, q)$ for $q \in \{.6, .9, .999\}$

An implication of the proposition is that, for any signal quality $p < 1$, a necessary condition for the existence of a full information equivalent separating debate equilibrium is that not only must there exist some informed committee members but that there must also be at least some uninformed individuals. To see the intuition here, let $(\mu, \upsilon)$ be a full information equivalent separating equilibrium and fix $p < 1$; then $\lambda < l_1(p)$. Because $\lambda < l_1(p)$ and $\mu$ is separating in common interest, there is no difficulty satisfying the pivotal voting constraints; it is the pivotal signaling constraints that bind. Specifically, from the proof to Proposition 2, it is the pivotal signal constraint on the uninformed individuals that defines those $(\lambda, p, q)$

for which the full information equivalent separating debate equilibria exist: any such triple must satisfy

$$\lambda \leq \frac{q(1-q)(2p-1)}{[(1-q)^2 + 2q^2(1-p)p + 2qp(1-q)]}.$$

Inspection of the constraint on $\lambda$ makes clear that as $q$ goes to either extreme for any information quality $p > 1/2$, an uninformed individual $i$ with $y$-bias becomes unwilling to offer a truthful speech in debate, thus upsetting the equilibrium.

There are three events where $i$'s message is pivotal: (a) both $j$ and $k$ are uninformed, have a bias for $x$, and send messages $m_j = m_k = 0$, or (b) both $j$ and $k$ are informed, have a bias for $x$, and send messages $m_j = -m_k = 1$, or (c) $j$ is uninformed and sends $m_j = s_j = 0$, $k$ is informed and sends message $m_k = s_k = 1$, and both $j$ and $k$ have a bias for $y$. If either event (a) or (b) occurs, $i$'s preferred outcome is $y$ and this is the committee decision if and only if $i$ sends the message $m'_i = -1$ rather than the truthful message $m_i = 0$. On the other hand, $i$'s most preferred outcome at event (c) is $x$ and this is the committee decision if and only if $i$ sends message $m_i = 0$. The critical pivotal event as $q$ goes to one is (b), the case in which both of the other committee members are almost surely informed but with opposing signals. As the probability of being uninformed becomes negligible, the likelihood of event (b) being true conditional on $i$ being pivotal increases in relative importance to the point that $i$ chooses to deviate from reporting her lack of information (inducing all individuals to vote their bias for $x$) in favour of influencing the committee to support $y$.[12] Similarly, when $q$ goes to zero the most likely signal pivot event is (a) with both $j$ and $k$ being uninformed; in this case $i$ believes that the committee decision depends almost surely on the distribution of bias in the case $i$ reports $m_i = s_i = 0$ but is (conditional on the event (a)) surely $y$ if she sends message $m'_i = -1$.

An alternative perspective on the separating equilibria identified in Proposition 2 is useful. Fixing $\lambda = 1/10$, Figure 3 identifies the set of parameter values $(p, q)$ for which there is a full information equivalent separating debate equilibrium. As $\lambda \to 1/2$, this region shrinks toward a neighbourhood of the point $(1, 1)$ and, as $\lambda \to 0$, the region expands to fill the space of all information structures. Loosely speaking, given any probability of being informed, $q$, high $p$ is equivalent to low $\lambda$.For $\lambda \leq l_1$ and signal strategy separating in common interest, the vote pivotal constraints are surely satisfied by individuals' voting on the basis of their full information induced preferences; the binding constraints here, therefore, are signal pivotal constraints. Perhaps surprisingly, the binding signal pivot constraint is not that insuring an individual with signal against her bias nevertheless finds it optimal to reveal that signal. Rather, the boundary of the region in Figure 3 for which the relevant equilibria exist is the set of informational structures at which uninformed individuals are just indifferent between revealing their lack of information and making a speech in support of their private interests.

[12] Of course, at $p$ and $q$ sufficiently large, the overall probability of an uninformed committee member being signal pivotal is negligible.

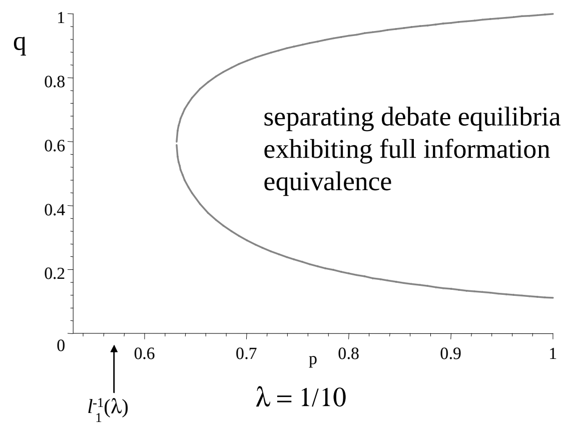


Figure 3: Separating debate equilibria

Of course, because making the latter speech is designed to encourage others to vote for the speaker's bias by influencing beliefs, the speech itself is not in terms of the speaker's bias *per se* but rather in terms of the common interest.

Proposition 2 claims that when individuals value the common good sufficiently highly and only a minimal amount of evidence in favour of an alternative being more in the common good is required to induce an individual to support that alternative, all private information on the common good can be credibly revealed in equilibrium and the subsequent voting behaviour results in full information equivalence. None of these properties, however, necessarily hold for semi-pooling debate equilibria.

## 4.2 Semi-pooling debate equilibria

By definition, in semi-pooling (SP) debate equilibria relatively informed individuals – that is, those for whom $s_i \neq 0$ – continue to make speeches advocating the alternative supported by their information, irrespective of their private interests, but uninformed individuals – those for whom $s_i = 0$ – now make speeches advocating the alternative they favour on private interests alone. In view of Proposition 1, therefore, speeches in SP debate equilibria involve everyone effectively announcing how they would have voted without debate.

Such speeches include no reference to private interests, but are indistinguishable in content from a speech given by an informed committee member; for example, $\mu(x,0,\lambda) = \mu(z,1,\lambda)$. As remarked earlier, while an informed agent may try to distinguish herself by arguing, for instance, that "While my private interests are for $y$ over $x$, nevertheless I currently believe $x$ is more likely to be in the common interest", such efforts at credibility fall on deaf ears; if such a speech is believed, an uninformed individual could make the same speech and be persuasive. Consequently, although it remains the case here that private interest oriented arguments have no impact in rational deliberation, the incentive for uninformed speakers to offer self-serving speeches about the common good necessarily leads listeners' interpretations of such speeches to be coloured by their beliefs about the likely private interests of the speaker, beliefs that the speaker can change only by changing his or her message regarding the common good. With SP debate equilibria, then, deliberation might improve the extent to which the committee achieves informational consensus but cannot insure much consensus in induced preferences.[13]

Beliefs regarding a speaker's private bias are not the only thing that distinguish interpretation of speech under SP from that under separating signaling strategies. In a separating debate equilibrium, the particular values of the parameters $q$ and $p$ play an important part in defining when full information equivalent voting constitutes equilibrium behaviour, but have nothing to do with the interpretation of debate speeches *per se*.[14] In SP debate equilibria, however, this is no longer true: the likelihood of being informed and the quality of any information in fact received bear both on the interpretation of speech and on subsequent voting decisions. Not surprisingly therefore, there can be a variety of SP debate equilibria that, at least observationally, differ exclusively in voting behaviour. Depending on the information structure, identical debates (that is, any $\mathbf{m} \in \mathcal{M}^3_\mu$ or permutation thereof ) can influence different individuals in different ways and lead to various profiles of voting decisions.

The different sorts of symmetric SP debate equilibria identified reflect different degrees to which individuals can be influenced by debate and signals. Table 1 describes the voting strategies, $\upsilon$, for the (symmetric) SP debate equilibria that exist. As indicated, each column headed by a bold-faced letter is a particular SP debate equilibrium and the voting behaviour is described in terms of an individual's signal, $s_i$, and the sum of the others' debate messages,

[13]Nor can SP equilibria in general insure justificatory consensus. By definition of an SP signal strategy, there are necessarily some decision-relevant facts left unsaid in debate (*viz.* whether a speaker is informed or not) and therefore some out-of-equilibrium messages that, depending on how individuals form the appropriate beliefs, could induce a change in expected outcome. In the analysis here, we adopt an explicit assumption about such out-of-equilibrium beliefs which yields justificatory consensus by fiat (all messages are assigned one of two meanings); but adopting a different refinement might induce different equilibria, upsetting any claim of justificatory consensus.

[14]Of course, $\lambda \leq l_1(p)$ is necessary for full information equivalent separating debate equilibria to exist at all. Nevertheless, any interpretation of (equilibrium) speech conditional on this constraint not binding is independent of $p$.

$M_{-i}$; the $y$-biased individual's prescribed vote is listed first.

| | | SP debate equilibria, $[v(y, s_i, \mathbf{m}), v(x, s_i, \mathbf{m})]$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $s_i$ | $M_{-i}$ | **A1** | **A2** | **B** | **C1** | **C2** | **R** | **M** | **U** |
| **-1** | -2 | y,y | y,y | y,y | y,y | y,y | y,y | y,y | y,x |
| | 0 | y,y | y,y | y,y | y,y | y,y | y,y | y,x | y,x |
| | 2 | y,y | y,y, | y,x | y,x | x,x | y,x | y,x | y,x |
| **0** | -2 | y,x | y,y | y,x | y,y | y,y | y,y | y,y | y,x |
| | 0 | y,x | y,x | y,x | y,x | y,x | x,y | y,x | y,x |
| | 2 | y,x | x,x | y,x | x,x | x,x | x,x | x,x | y,x |
| **1** | -2 | x,x | x,x | y,x | y,x | y,y | y,x | y,x | y,x |
| | 0 | x,x | x,x | x,x | x,x | x,x | x,x | y,x | y,x |
| | 2 | x,x | x,x | x,x | x,x | x,x | x,x | x,x | y,x |

Table 1: SP debate equilibrium voting strategies under majority rule

Some language is useful. Given a signaling strategy $\mu$, an individual $i \in N$ with signal $s_i$ *can be influenced in debate* if $i$'s vote choice is not constant on $\mathcal{M}^3_\mu$; that is, there exist distinct $\mathbf{m}, \mathbf{m}' \in \mathcal{M}^3_\mu$ such that $v(b_i, s_i, \mathbf{m}) \neq v(b_i, s_i, \mathbf{m}')$. On the other hand, say that $i$ *surely votes her signal* if $i$ votes her signal whatever other speeches are offered in debate. Then referring to Table 1:

∘ SP equilibria A1 and A2 have all informed individuals surely vote their signals and, although the uninformed always vote their bias in A1, they can be influenced in debate under A2.

∘ By contrast, in B it is only those informed individuals with a signal *against* their bias who can be influenced in debate, albeit minimally by two speeches *that favour* their bias; all other informed types vote their signal and the uninformed vote their bias.

∘ SP equilibrium C1 integrates the informed agents' voting behaviour under B with the uninformed agents' voting behaviour under A2; that is, both the uninformed and (minimally) some informed individuals can be influenced in debate.

∘ C2 is the 'most influential' of the SP equilibria available; along the equilibrium path, every individual's voting behaviour in C2 coincides with that in a separating debate equilibrium. Indeed, this last property is true of all uninformed voters in C2, C1 and A2 SP equilibria.

∘ In the equilibrium R, informed individuals are influenced in debate exactly as in B and C1; the singular feature of R is the voting behaviour of the uninformed. Although they make speeches in support of their bias, they vote *against* their bias unless (like the informed individuals) they hear two speeches in debate that favour their bias. To develop some intuition for why such voting behaviour by uninformed individuals may not be absurd, recall that under semi-pooling debate uninformed individuals speak in favour of their bias. So, for instance, conditional on hearing a split debate $(m_j, m_k) = (1, -1)$, an uninformed individual $i$ might reason that if her vote is pivotal, it is most likely to be the speaker who presented the minority opinion, say $j$, who is voting against the majority position advocated in debate. In turn, this suggests that $j$ is relatively more likely to be informed in which case, conditional on being pivotal, $i$ voting for the minority deliberative opinion is the best thing to do.

∘ Finally, debate in both equilibria M and U has very little impact. Although all individuals can be influenced by debate in M, every committee member votes their bias unless they listen to two speeches against that bias *and* they have no private signal for their bias and, in U, all individuals surely vote their bias irrespective of their private signal or any debate.

Rather than state a long and tedious proposition delineating the formal conditions on triples $(\lambda, p, q)$ for which each SP debate equilibrium exists, we describe things graphically with a canonical example, setting $\lambda = 1/10$: see Figure 4, on which the boundary for the full information equivalent separating equilibrium is superimposed (the dotted outline).[15] Before discussing the diagram in any detail, however, it is worth remarking that, whereas the binding constraint for the separating debate equilibrium is at the debate stage through the pivotal signaling constraints, with exception of the south-west boundary of the C1 SP debate equilibrium, the binding constraints on the SP equilibria are at the voting stage, through the pivotal voting constraints. Insofar as debate in SP equilibria is constant and only the voting responses to debate changes with the information structure, this shift in which constraints bite has some intuition. What is perhaps somewhat less intuitive, is that there are frequently discontinuities in these constraints as we move from one SP debate equilibrium to another (for instance, between A1 and A2 in Figure 4); such discontinuities arise because even marginal changes in voting responses to given debates induce non-marginal shifts in the identity and conditional likelihoods of voting pivot events. Such subtleties yield considerable complexity.To give some intuition about what is happening here, imagine the set of feasible information structures $(\frac{1}{2}, 1) \times (0, 1)$ divided loosely into four subsets, according to whether the quality of information, $p$, and the probability of being informed, $q$, are “low” or “high”.

When both $p$ and $q$ are low, debate has very little impact on subsequent voting behaviour. At the extreme, for sufficiently small $p$ and $q$, both signals and debate are quite irrelevant (U); at somewhat higher values of $p$ relative to $q$, private signals become influential but debate remains uninfluential (A1). In this case, low $q$ implies the reliable informational content of any set of speeches is insufficient to offset any bias among the uninformed or constitute an effective counter to the direct influence of a private signal. Increasing the likelihood of being

[15] The derivations supporting Figure 4 are sketched in Appendix B.

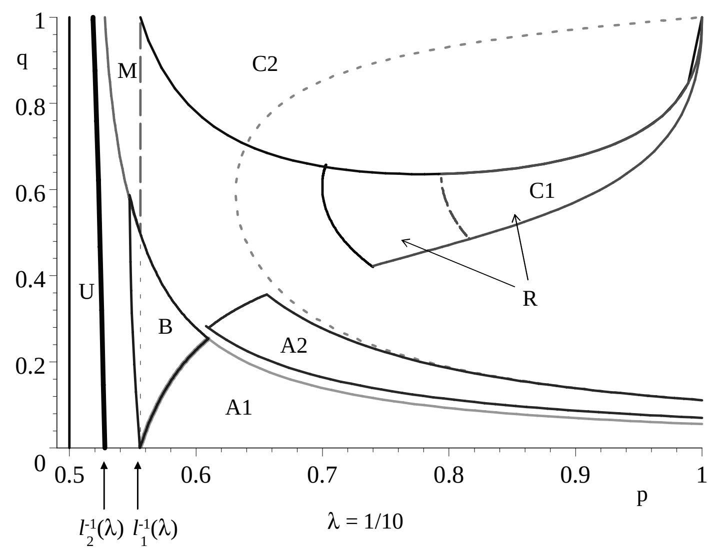


Figure 4: Semi-pooling debate equilibria

informed ($q$), however, can result in the uninformed becoming subject to influence in debate (A2) although, under both A1 and A2 equilibria, the informed surely vote their signals. Thus deliberation can influence two persons with the same bias in different ways. Alternatively, keeping $p$ low and increasing $q$ relative to $p$ leads to SP debate equilibrium B, where the uninformed vote their bias and only the informed can be influenced by debate. However, the extent to which the informed are persuadable is very limited, being restricted to those with private information against their bias voting against this information on hearing two arguments for their bias.

In sum, at low $p$/low $q$ information structures, the generally poor quality and thin distribution of information on common interests among the committee leads to an extremely limited role for deliberation in majority decision making. The observation applies fairly directly to high $p$/low $q$ information structures. To all intents and purposes, the only change here when $q$ is very small is that there is no chance for the informed to be influenced in debate: SP equilibrium B is unavailable. On the other hand, at high $p$ full information equivalent separating equilibria can be supported at relatively low (but not too low) values of $q$ as compared with low $p$ environments.

Now consider information structures in which the likelihood of being informed is high but the quality of private signals is relatively low. If $p$ is sufficiently small, again we find debate having very little impact: either voting is wholly unresponsive to informative debate (U) or it is barely responsive (M). On the other hand, for somewhat higher signal quality it is possible to support the C2 equilibrium, where debate is as informative and influential as it can be under semi-pooling signaling in common interests. Although all individuals' voting decisions in C2 are precisely those consistent with treating all speeches as true, not all of the decision-relevant information is offered in semi-pooling debate (the uninformed never indicate this fact) and so full information equivalence is not assured. For example, let $(\mu, \upsilon)$ be a C2 semi-pooling debate equilibrium; suppose $\mathbf{s} = (1, 0, 0)$ and individuals 2 and 3 are $y$-biased. Then by definition of $\mu$, the debate is $\mathbf{m} = (1, -1, -1)$ and the C2 voting strategy $\upsilon$ yields a unanimous vote for $y$. But C2 equilibria only exist for $\lambda \leq l_1(p)$ which implies the right committee decision at $\mathbf{s}$ is $x$.

Finally, suppose both $p$ and $q$ are high. Then the SP equilibria are C1, R and C2, although C1 equilibria exist only if R equilibria exist and R equilibria exist only if the separating debate equilibrium exists. Informed individuals respond identically to debate in C1 and R but differently in C2. The difference is in the decisions of those with private information supporting their bias: under C1 such an individual could not be influenced in debate whereas under C2 she votes as she would under the separating signaling strategy. On the other hand, uninformed individuals respond identically to debate in C1 and C2 but differently in R. Indeed, their response to debate in R is exactly the reverse of that in C1 and C2: in C1 and C2, the uninformed vote their bias unless they hear two speeches against it; in R the uninformed vote against their bias unless they hear two speeches for it.

There are two further things to note about C2 equilibria. First, although voting behaviour here is "sincere" in that it reflects the balance of bias, signal and debate for each committee member, individual debate behaviour is not so sincere; by definition of a semi-pooling signaling strategy, the uninformed misrepresent their knowledge in debate, arguing for their bias by adopting the speech of those informed in favour of that bias. The second, related thing, to note is that despite the fact that occasionally some speeches may not reflect any private information, such strategic speech-making turns out not to lead to any worse an outcome than the one under no debate. To see this, consider the following example.

Let $(p, q)$ be an information structure for which both separating and C2 semi-pooling debate equilibria exist. Suppose individuals $i = 1, 2$ are two uninformed individuals with a $y$-bias and assume individual $i = 3$ is informed with a signal $s_3 = 1$ (her bias is irrelevant here). Then under SP signaling, the equilibrium debate is $\mathbf{m} = (-1, -1, 1)$ whereas the realized signal profile is $\mathbf{s} = (0, 0, 1)$. Under C2 voting, therefore, all individuals vote for $y$ and $y$ is chosen; under separating signaling, however, $\mathbf{m}' = (0, 0, 1)$ and all individuals vote for $x$ and $x$ is chosen. Clearly, the change in induced preferences under C2 supports a wrong committee decision. On the other hand, in the absence of debate, the unique equilibrium voting profile has individuals 1 and 2 both vote $y$ and individual 3 vote for $x$, so again $x$ is chosen.

The preceding example is not an artifact. Before making this assertion precise, it is useful to check some intuitive properties of voting behaviour in any (anonymous although not necessarily symmetric) SP debate equilibrium under majority rule.

**Lemma 1** *In any SP debate equilibrium under majority rule,*

$$\upsilon(y, 0, -1, -1, -1) = y.$$

Assuming the committee makes decisions under majority rule, therefore, an uninformed $y$-biased agent surely votes her bias following any (semi-pooling) debate in which everyone argues for choosing $y$ over $x$ and, evidently, a completely symmetric argument applies for $x$-biased individuals; that is, $\upsilon(x, 0, 1, 1, 1) = x$ also. In other words, the symmetry of majority rule coupled with that of the semi-pooling message strategy $\mu$ implies a considerable degree of symmetry in SP equilibrium voting.[16]

The next result, Lemma 2, says that SP debate equilibrium vote decisions are signal (claim 1) and bias (claim 2) monotonic.

**Lemma 2** *Let $(\mu, \upsilon)$ be any SP debate equilibrium and $\mathbf{m} \in \mathcal{M}^3_\mu$. Then, under majority rule, for all signals $s > s'$ and each bias $b$:*

$$(1) \begin{cases} [\upsilon(b, s, \mathbf{m}) = y \Rightarrow \upsilon(b, s', \mathbf{m}) = y], \\ [\upsilon(b, s', \mathbf{m}) = x \Rightarrow \upsilon(b, s, \mathbf{m}) = x]; \end{cases}$$

$$(2) \begin{cases} [\upsilon(x, s, \mathbf{m}) = y \Rightarrow \upsilon(y, s, \mathbf{m}) = y], \\ [\upsilon(y, s, \mathbf{m}) = x \Rightarrow \upsilon(x, s, \mathbf{m}) = x]. \end{cases}$$

It seems sensible that in addition to bias and signal monotonicity, debate equilibria should also exhibit some sort of monotonicity in messages: if, given a distribution of bias and information in the committee, the only difference between two debates $(m_i, \mathbf{m}_{-i})$, $(m_i, \mathbf{m}'_{-i})$ from $i$'s perspective is that the speeches of others $\mathbf{m}'_{-i}$ are both more favourable to the individual's bias than are $\mathbf{m}_{-i}$, then $i$ should vote his bias following $(m_i, \mathbf{m}'_{-i})$ if $i$ votes his bias following $(m_i, \mathbf{m}_{-i})$. This sort of monotonicity is satisfied by all of the debate equilibria considered so far and, as will become apparent shortly, all of those discussed in the next section on unanimity rule. But messages and votes are strategic decisions and, at least as far as we know at present, this form of monotonicity is not implied by the current assumptions on equilibrium behaviour.

**Definition 2** *A voting strategy $\upsilon$ satisfies debate monotonicity if and only if, for all $(b_i, s_i, m_i)$, $\upsilon(b_i, s_i, m_i, m_j, m_k) = b_i$, $b_i m'_j \geq b_i m_j$ and $b_i m'_k \geq b_i m_k$ imply $\upsilon_i(b_i, s_i, m_i, m'_j, m'_k) = b_i$.*

[16] It is worth remarking that such symmetry cannot be derived when there is no debate. Majority rule by itself is insufficient to rule out the possibility, say, of two informed individuals voting differently with the same signal.

In words, if an agent is voting consistent with his bias after observing a signal and some debate then he must also be voting for his bias if he observes the same signal (and therefore sends the same message) and a debate that is more favourable for his bias. Note that debate monotonicity requires holding constant the agent's bias, signal and message. Moreover, debate monotonicity does not imply, for instance, that an uninformed $y$-biased individual who sends a message $m = -1$ and hears a split debate $\mathbf{m}_{-i} = (1, -1)$ surely votes his bias. Requiring debate monotonicity, then, is a substantively weak restriction; it is nevertheless very useful analytically.

**Lemma 3** *In any symmetric SP debate equilibrium in which voting is debate monotonic, either there is a positive probability the vote of agent $i$ is pivotal given debate $(m_i, m_j, m_k) = (1, -1, -1)$ or agents $j$ and $k$ are both voting for $y$.*

Recall the definition of a "right" committee decision as being the decision reached under decision making with full information on $\mathbf{s}$; a "wrong" committee decision is any decision that is not "right". There are, therefore, two sorts of error in committee decision making: either, irrespective of bias, induced preferences are unanimous at $\mathbf{s}$ and there is an *error in common interests*, or the distribution of induced preferences at $\mathbf{s}$ coincides exactly with distribution of bias and there is an *error in bias*.

Let $\alpha$ and $\beta$ be two institutional forms of committee decision making. Say that $\alpha$ *weakly dominates $\beta$ at $(\lambda, p, q)$ with respect to common interest (bias)* if (1) whenever $\alpha$ yields an error in common interest (bias) then $\beta$ also yields an error in common interest (bias); and (2) $\beta$ sometimes yields an error in common interest (bias) when $\alpha$ yields the right decision. *$\alpha$ weakly dominates $\beta$ at $(\lambda, p, q)$* if $\alpha$ weakly dominates $\beta$ at $(\lambda, p, q)$ with respect to both common interest and bias.

**Proposition 3** *Assume only symmetric pure strategy debate equilibria are played and that committee decisions are made by majority rule. If equilibrium voting is debate monotonic then, with respect to common interest, committee decision making with debate weakly dominates committee decision making without debate at almost all $(\lambda, p, q)$.*

Two things are worth emphasizing about Proposition 3, the main result of this section. First, the result does not refer only to those SP debate equilibria identified in Figure 4, but applies quite generally to all symmetric pure strategy SP debate equilibria exhibiting debate monotonicity (both with and without the technical refinement); and second, the result does not say that for every feasible $(\lambda, p, q)$ there exists a debate equilibria that is, with respect to yielding "right" decisions at $(\lambda, p, q)$, weakly better with respect to errors in common interests than the no-debate equilibrium, but rather that *every* symmetric pure strategy debate equilibrium has this property at any feasible $(\lambda, p, q)$.

Proposition 3 does not extend to errors in bias that are not also errors in common interests. Assuming $y$ is the right decision at some situation $\mathbf{s}$, such an error in bias can only occur if

(up to permutations)

$$\mathbf{s} \in \{(0,0,0), (0,1,-1)\},$$

implying the probability that $Y$ is the true state is 1/2. *A priori*, therefore, there seems little reason to think that such errors in bias are any less important than errors in common interest. And although majority rule without debate is not immune to errors in bias (an example is given below), it is possible for debate to yield such an error where none would be made in its absence.

To see the difficulty, suppose $\mathbf{s} = (0,0,0)$ with $y$ the right decision; this implies that at least two of the committee are $y$-biased (say, $i = 1, 2$), so $y$ is surely the no-debate equilibrium decision (Proposition 1). Now let $(\mu, \upsilon)$ be a symmetric semi-pooling debate equilibrium satisfying debate monotonicity. There are two debates possible in equilibrium, depending on individual 3's bias. If $b_3 = y$ then $\mathbf{m} = (-1,-1,-1)$ and Lemma 1 implies there is no error; so suppose $b_3 = x$, yielding $\mathbf{m} = (-1,-1,1)$. By anonymity, if $(\mu, \upsilon)$ results in an error at this debate, each uninformed $y$-biased individual $i \in \{1,2\}$ must vote for $x$ conditional on sending a message $m_i = -1$ and hearing a split debate $\mathbf{m}_{-i} = (-1, 1)$; that is an error in bias implies

$$\upsilon(y, 0, -1, -1, 1) = x.$$

But this exactly describes the voting behaviour of uninformed types in the SP debate equilibrium R. Similar reasoning applies to the remaining possibility, $\mathbf{s} = (0, 1, -1)$: if the bias distribution is $\mathbf{b} = (y, x, y)$, then the debate has to be $\mathbf{m} = (-1, 1, -1)$ and, in the SP debate equilibrium R, the first two individuals vote for $x$ to produce an error. In this case, however, there is no guarantee that the no-debate equilibrium outcome is right: if $\mathbf{b}' = (x, y, y)$, the right decision is $y$ but the no-debate outcome is $x$.

It is not hard to see from the preceding discussion that a necessary condition for an error in bias (not involving errors in common interest) to result from debate is that uniinformed individuals vote against their bias conditional on observing a split debate. In Appendix A, we show that the only symmetric SP debate equilibrium exhibiting such behaviour is R and R exists only if the separating debate equilibrium also exists. This fact immediately implies the following simple corollary

**Corollary 1** *Assume the separating debate equilibrium is surely played whenever it exists. Then, under the hypotheses of Proposition 3, committee decision making with debate weakly dominates committee decision making without debate at almost all* $(\lambda, p, q)$.

## 5 Unanimity rule

Insofar as debate results in some form of preference consensus, as claimed or postulated in much of the normative literature on deliberation, then any final committee decision by voting will be unanimous whatever the formal institutional rules prescribe. If only informational or justificatory consensus emerges from conversation then, as explicitly recognized in the earlier quote from Cohen, unanimity in voting is unlikely and the voting stage constitutes more than a formal ratification of the consensual decision. It seems sensible, then, to ask how variations in the voting rule might influence the character and extent of deliberation. Furthermore, there is some intuition that requiring unanimity at the voting stage promotes more information sharing and argument in any debate, since some form of consensus is now essential for pro-active committee decision.

Unlike with majority rule, unanimity rule requires a default choice or status quo. Without loss of generality, then, suppose $x$ is the status quo and can be rejected in favour of $y$ only if all three committee members vote for $y$ against $x$.[17] And since unanimity rule is evidently not symmetric, there is no good reason to insist, or even focus, on symmetric equilibria; in fact, quite the contrary is true. Consequently, we no longer look for symmetric voting strategies, although we maintain the presumption of anonymity.

Suppose there is no debate stage and note that an individual is pivotal in voting only if both of the other committee members are voting for $y$. Then it cannot be the case that all types surely vote for $y$ in any equilibrium, irrespective of their bias or signal. It is easy to see why: suppose the claim false and consider an individual $i$ with signal $s_i = 1$ and bias for $x$. Then the event that $i$ is vote pivotal under unanimity contains no additional decision-relevant information for $i$, in which case, given signal $s_i = 1$, voting for $x$ surely is the best decision. In fact, it turns out that Table 2 describes the only pure strategy profiles that can constitute no-debate equilibria. As for Table 1 above, each column headed by a bold-faced letter is a particular pure strategy (no-debate) voting equilibrium; because there is no debate, any individual's vote strategy can be described in terms of the individual's signal, $s_i$. And again, the $y$-biased individual's prescribed vote is listed first.

| | No-debate equilibria, $[v(y, s_i), v(x, s_i)]$ | | | | |
|---|---|---|---|---|---|
| $s_i$ | **1** | **2** | **3** | **4** | **5** |
| **-1** | y,y | y,y | y,y | y,y | y,x |
| **0** | y,y | y,y | y,x | y,x | y,x |
| **1** | x,x | y,x | x,x | y,x | y,x |

Table 2: No-debate voting strategies under unanimity

In stark contrast to the situation with majority decision, no-debate equilibrium voting under unanimity is quite complicated; in particular, for some information structures equilibrium existence requires mixed strategies. With this in mind, let **a-b** denote a mixed voting strategy profile that involves individuals randomizing between their respective vote decisions

[17] A seemingly plausible alternative assumption here, is to take a fair lottery over $\{x, y\}$ as the status quo and require a unanimous vote to insure either alternative surely. But then decision making is over a three, rather than two, alternatives, a quite different scenario.

under pure strategies **a** and **b**, where **a**,**b** take values in $\{\mathbf{1,\ldots,5}\}$. Then Figure 5 below describes the distribution of voting equilibria under unanimity rule with no debate, assuming $\lambda = 1/10$; the value of $\lambda$ here is purely one of convenience and the diagram is canonical. And note that in those regions with mixed strategy equilibria, exactly one type of person is ever required to use a non-degenerate lottery; for instance, in the **1-2** equilibrium, only $y$-biased individuals with signal $s = 1$ are required to randomize.

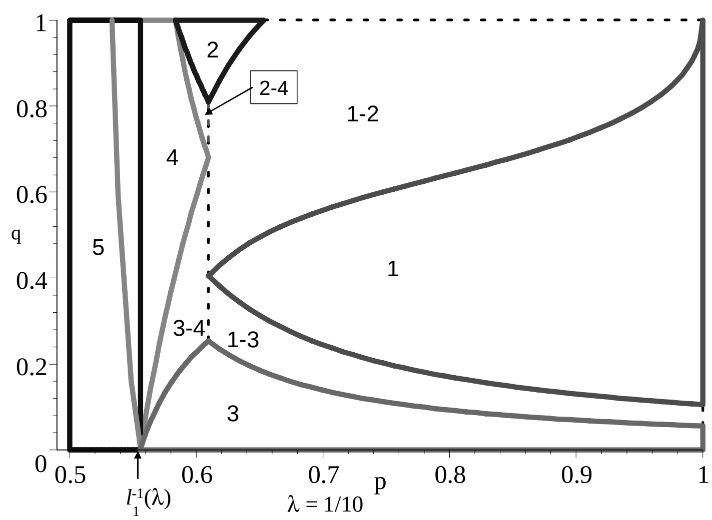


Figure 5: Voting equilibria under unanimity with no debate

In common with the story for SP debate equilibrium voting under majority rule, broadly speaking, the better the quality of the information the more willing are uninformed individuals to vote for $y$, effectively delegating the committee decision to the informed committee members, who likewise are more willing to vote their signal irrespective of bias (**1**). And again, the presence of uninformed individuals is important here. As signal quality declines, individuals with an $x$-bias become increasingly unwilling to vote against their bias.

Despite the fact that the voting strategies described in Figure 4 are predicated on SP debate equilibria under majority rule, whereas those described in Figure 5 concern no-debate equilibria under unanimity rule where the pivot events are quite different, there are some close similarities between the regions delineated in the two figures. For example, the pivotal voting constraints defining the SP debate equilibrium A1 under majority voting are identical to those defining the no-debate voting equilibrium **3** under unanimity rule. A partial explanation for this coincidence can be found by noting that, conditioning on the event $M_{-i} = 0$, the two voting strategies coincide. That this is not the whole story can be seen by comparing the region in Figure 5 where the no-debate equilibrium **5** with unanimity exists, with the region in Figure 4 where SP equilibrium U can be found. Although very similar, these regions are not identical; nevertheless, conditional on the event $M_{-i} = 0$, the two voting strategies are identical.

The idea of a "right" committee decision adopted for majority rule was in terms of majority preference conditional on common knowledge of $\mathbf{s}$. Moreover, we remarked that, at least for $\lambda < l_1$, this definition coincides with an alternative notion, that a right decision is an alternative most likely in the common interest conditional on common knowledge of $\mathbf{s}$. Exactly the same observation applies for unanimity rule: for $\lambda < l_1$, an alternative preferred by all individuals conditional on common knowledge of $\mathbf{s}$ is also an alternative most likely in the common interest conditional on common knowledge of $\mathbf{s}$. This justifies leaving the concept of a "right" committee decision unchanged, despite the change in voting rule from majority to unanimity.

It is immediate that no-debate equilibrium under unanimity can yield errors in bias: suppose the equilibrium is **1** and all individuals are both uninformed and $x$-biased; then there is a unanimous vote for $y$ where in fact $x$ is the right decision.

The multiplicity of no-debate equilibria under unanimity rule makes an unequivocal statement about the likelihood of an equilibrium committee decision being "right" contingent on the particular equilibrium played. However, the bounds are clear. The smallest likelihood of error is when voting strategy **1** is equilibrium behaviour. Here, an error occurs only in state $Y$ when all individuals are informed but one sees an incorrect signal; this occurs with probability $q^3p^2(1-p)/2$. At the other extreme for $\lambda \leq l_1$, when **4** is equilibrium behaviour there are multiple events at which error can occur; doing the (tedious) calculation gives the likelihood of error as $\left[4pq\left(1-2q+pq\right) - pq^3\left(p+2p^2-4\right)\right]/16$. And for $\lambda > l_1$, under **5**, the likelihood of error is $1/2$. More interesting, is what happens when deliberation precedes any committee vote.

**Proposition 4** *There exists no separating debate equilibrium under unanimity.*

Comparing this result with Proposition 2 undermines any general claim that requiring unanimity to make policy changes induces more deliberation in committee than requiring only a majority. Depending on individual attitudes toward the common interests and on the information structure, deliberation can be fully informative under majority rule but not under unanimity.[18]

[18] A similar impossibility result is proved by Doraszelski, Gerardi and Squintani (2001), the only other model that, to the best of our knowledge, considers deliberation under unanimity rule. DGS study a two-person committee that is choosing between a status quo and a given alternative policy; rejection of the status quo requires unanimous approval. There are two states of the world from the common interest perspective,

Although full separation, and thereby full information equivalence, is impossible under unanimity, deliberation can nevertheless be informative under some circumstances. We prove the claim by example. Under unanimity rule, any individual who so chooses can guarantee a committee decision for $x$ with her vote alone. This suggests a debate strategy in which a $y$-biased individual argues for $y$ in debate irrespective of any private information: if such an individual is persuaded, either by her private information or by deliberation, that $x$ is most in her interests then her own vote insures this outcome whatever she says in debate; but if she is left preferring $y$ over $x$ then her deliberative argument can be pivotal. Similarly, an individual for whom $x$ is most in his private interests has nothing to lose by sharing his information on the relative common good properties of the two alternatives. Formally, the suggested pattern of deliberation is described by the asymmetric message strategy, $\bar{\mu}$: for all distinct $s, s' \in \{-1, 0, 1\}$,

$$\bar{\mu}(x, s) \neq \bar{\mu}(x, s') \text{ and } \bar{\mu}(y, s) = \bar{\mu}(x, -1).$$

Thus, under $\bar{\mu}$, all $y$-biased individuals pool in common interests and all $x$-biased individuals separate in common interest. For want of a better term, then, call any debate equilibrium $(\bar{\mu}, \upsilon)$ a *bias-driven debate equilibrium.*

In any bias-driven debate equilibrium, those who are most likely to want change ($y$-biased individuals) argue consistently for this alternative irrespective of their signal, so suppressing any information they might have in support of the status quo $x$; against this obscurantism, those most likely to resist change ($x$-biased individuals) are willing to reveal all of their information in debate, whether or not it suggests that in fact $y$ is the better alternative on common interest grounds. But despite this willingness on the part of $x$-biased committee members to make a case for $y$ when appropriate, the only credible arguments are those who argue (at least weakly) on behalf of $x$; any effort by an $x$-bias individual to argue for $y$ is confounded by the incentives for those with a private interest for $y$ also arguing that case. So there is small hope here of achieving any sort of consensus through deliberation alone. But such a lack of deliberative consensus need not imply that deliberation cannot yield unanimous voting in committee.

Equilibria involving such asymmetric deliberation do exist; Figure 6 illustrates an example for $\lambda = 1/10$. As indicated in the diagram, a necessary but not sufficient condition on the

say $X$ and $Y$, and both individuals strictly prefer $x$ (respectively, $y$) in state $X$ (respectively, $Y$). Where they differ is in the attitudes about making errors and these attitudes (parametrized by some real number from the unit interval) are private information. In addition to learning their particular attitude to error, each individual also observes a noisy binary signal regarding the true state of the world. Inter alia, DGS study what happens when both individuals can give cheap-talk signals about their signals prior to voting. Their main results are that there is no separation in debate and deliberation is influential only in the case when an individual's signal conflicts with her disposition and prior belief: "When there is a conflict between a player's preferences and her private information about the state, she votes in accordance with her private information only if it is confirmed by the message she receives from her opponent" (p.2).

information structure $(p, q)$ for the bias-driven equilibrium to exist is that the no-debate voting equilibrium **1** also exists at $(p, q)$.

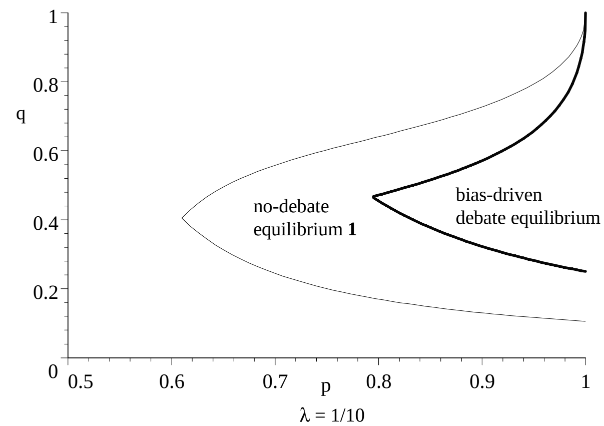


Figure 6: Bias-driven debate equilibrium $(\bar{\mu}, \bar{v})$

The identified bias-driven equilibrium is the pair $(\bar{\mu}, \bar{v})$. The message strategy $\bar{\mu}$ is defined above and the voting strategy $\bar{v}$ is described in Table 3, where the pairs in the two "$m_i$"-columns are the votes, $(\bar{v}(y, \cdot), \bar{v}(x, \cdot))$.

| $s_i$ | $M_{-i}$ | $m_i \in \{-1, 0\}$ | $m_i = 1$ |
|---|---|---|---|
| **-1** | $\leq 0$ | y,y | y,y |
| | 1 | y,x | y,x |
| | 2 | x,x | x,x |
| **0** | -2 | y,y | y,y |
| | -1 | y,y | y,y |
| | [-1+1] | **y,x** | **x,x** |
| | [0+0] | y,x | y,x |
| | 1 | x,x | x,x |
| | 2 | x,x | x,x |
| **1** | $\geq -2$ | x,x | x,x |

Table 3: Voting strategy $\bar{v}$

The binding constraints on $(\bar{\mu}, \bar{v})$ are two pivotal voting constraints: the lower boundary illustrated in Figure 6 describes the locus of information structures at which an $x$-biased individual with signal against his bias is indifferent at $\bar{v}$ between voting for $y$ (as required) or $x$ conditional on hearing a split debate, $(m_j, m_k) = (-1, 1)$; the upper boundary describes the locus of information structures at which an $y$-biased individual with signal against her bias is indifferent at $\bar{v}$ between voting for $x$ (as required) or $y$ conditional on hearing a uniform debate, $(m_j, m_k) = (-1, -1)$.

The most interesting thing to note about the strategy $\bar{v}$ is that an uninformed $y$-biased individual $i$ votes for $y$ against $x$ if $i$ makes any speech $m_i \in \{-1, 0\}$ (weakly) in support of choosing $y$, but votes for $x$ against $y$ if, for some reason, $i$ advocates choosing $x$, $m_i = 1$, *and* the others are divided in debate, $(m_j, m_k) = (-1, 1)$. In other words, under $\bar{v}$, an individual with a given signal, hearing given speeches by others in debate, nevertheless votes differently depending on the particular cheap talk speech she delivers; in this case, the individual "talks herself into voting against her bias". Such behaviour is not, it turns out, unreasonable: because the subsequent votes of others depend in part on the arguments they hear in debate, the pivotal voting constraint facing an individual following one speech does not necessarily coincide with that following a different speech. In fact, although, in the equilibrium $(\bar{\mu}, \bar{v})$, this particular behaviour is off-equilibrium-path, it proves essential to support existence of $(\bar{\mu}, \bar{v})$ as equilibrium behaviour at all. If, as seems intuitive, the uninformed $y$-biased individual's vote is independent of her own message at any debate (in particular, at the debate $(m_j, m_k) = (-1, 1)$), then not all of the signal pivotal constraints can be satisfied along the equilibrium path.

Similar considerations apply, although less evidently, elsewhere in the equilibrium voting profile. From Table 3, an $x$-biased individual with a signal against her bias $(s_i = -1)$ is required to vote for $y$ conditional on $M_{-i} = -1$ whatever speech she makes. However, if the probability of others being informed, $q$, is sufficiently low, then such an individual strictly prefers to vote for $x$ in the event she sends the off-path message $m_i' = 1$ supporting her private bias rather than her signal, $\bar{\mu}(x, -1) = -1$ (or a speech $m_i = 0$) but not otherwise. Moreover, if the individual is presumed to vote for $x$ conditional on sending $m_i' = 1$, then $(\bar{\mu}, \bar{v})$ cannot describe equilibrium behaviour at any information structure.

Recall that the probability of the committee choosing the wrong alternative in the no-debate equilibrium $\mathbf{1}$ is $q^3 p^2 (1-p)/2$. Under the bias-driven debate equilibrium $(\bar{\mu}, \bar{v})$, the probability of the committee making an error in common interest falls to zero but that of making an error in bias remains strictly positive: if $b_1 = y$, $b_2 = b_3 = x$ and all individuals are uninformed, the debate under $(\bar{\mu}, \bar{v})$ is $\mathbf{m} = (-1, 0, 0)$ and, given $\bar{v}$, all individuals vote for the wrong outcome, $y$, exactly as in the no-debate equilibrium $\mathbf{1}$. Nevertheless, it is clear by inspection that the debate equilibrium $(\bar{\mu}, \bar{v})$ also weakly dominates the no debate equilibrium $\mathbf{1}$ with respect to bias $(\bar{\mu}, \bar{v})$, that is, when $\mathbf{s} \in \{(0,0,0), (0,-1,1)\}$. Whenever there is an error in bias alone under $(\bar{\mu}, \bar{v})$ there is also an error under $\mathbf{1}$ without debate, but the converse is false: let $\mathbf{s} = (0, 0, 0)$ and $b_i = x$ all $i$; then without debate the wrong decision $y$ is made but with debate the decision is $x$. It follows that the bias-driven debate equilibrium $(\bar{\mu}, \bar{v})$ weakly dominates the no debate equilibrium $\mathbf{1}$. This is perhaps to be expected: debates supported by $\bar{\mu}$ necessarily make committee members strictly more informed at the voting stage than they are without debate.[19] In general, however, the weak dominance result for $(\bar{\mu}, \bar{\upsilon})$ does not extend to all bias-driven debate equilibria.

**Proposition 5** *Assume committee decisions are made by unanimity rule. There exist* $(\lambda, p, q)$ *at which the committee makes the wrong decision under a bias-driven debate equilibrium* $(\bar{\mu}, \hat{\upsilon})$*, but makes the right decision under a no-debate equilibrium,* $\upsilon^0$*.*

**Proof** We show by example that bias-driven debate can support errors in common interest in settings where the committee decision under the relevant (pure strategy) no-debate equilibrium is the right decision. Assume $\lambda = 1/10$ (this particular value is inessential). The message strategy $\bar{\mu}$ is defined above; the vote strategy $\hat{\upsilon}$ is described in Table 4 where, as usual, the "$m_i$"-columns are the votes, $(\hat{\upsilon}(y, \cdot), \hat{\upsilon}(x, \cdot))$.

| $s_i$ | $M_{-i}$ | $m_i = -1$ | $m_i = 0$ | $m_i = 1$ |
|---|---|---|---|---|
| **-1** | $\leq -1$ | y,y | y,y | y,y |
| | [-1+1] | y,y | y,y | y,x |
| | [0+0] | y,y | y,y | y,y |
| | 1 | y,x | y,x | y,x |
| | 2 | x,x | x,x | x,x |
| **0** | -2 | y,y | y,y | y,y |
| | -1 | y,x | y,x | y,y |
| | [-1+1] | y,x | y,x | y,y |
| | [0+0] | y,x | y,x | x,x |
| | 1 | x,x | x,x | x,x |
| | 2 | x,x | x,x | x,x |
| **1** | $\geq -2$ | x,x | x,x | x,x |

Table 4: Voting strategy $\hat{\upsilon}$

Insisting on the technical equilibrium refinement (individually independent trembles) leads to difficulties off the (postulated) equilibrium path here. In particular, the strategy pair $(\bar{\mu}, \hat{\upsilon})$ is an equilibrium and survives the refinement only on a line cutting through the set $(p, q)[\mathbf{1}] \subset (\frac{1}{2}, 1) \times (0, 1)$ on which the no-debate voting equilibrium **1** exists. However, the no-debate equilibrium **1** obviously exists without insisting on the refinement and lifting the refinement further results in $(\bar{\mu}, \hat{\upsilon})$ constituting equilibrium behaviour on a nonempty set of information structures having strictly positive measure: see Figure 7.[20]In the figure, the

[19] This is true even if the debate is $\mathbf{m} = (-1, -1, -1)$; in this case all individuals know there is no $x$-biased committee member with a signal $s \geq 0$.

[20] As indicated in Appendix B, establishing these claims formally is both tedious and computationally demanding, so we omit the details. All of the derivations supporting this example and the figures in the text, however, are available from the authors on request.

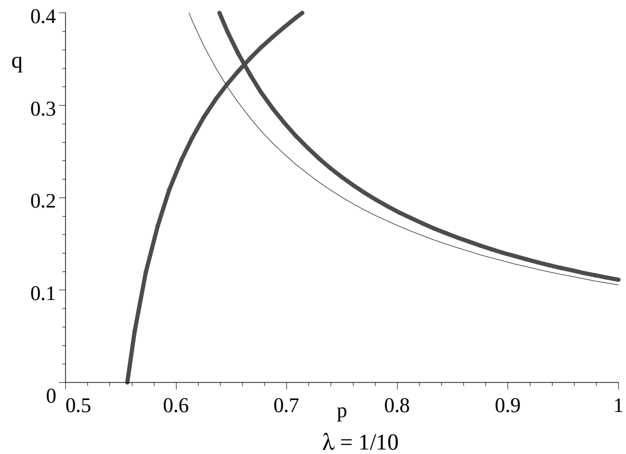


Figure 7: Bias-driven debate equilibrium $(\bar{\mu}, \hat{v})$

region below the two intersecting thick lines is the set of information structures for which $(\bar{\mu}, \hat{v})$ is an (unrefined) equilibrium; the downward sloping thin line is the lower boundary of $(p, q)[\mathbf{1}]$. If the technical equilibrium refinement is imposed, the set of information structures delineating those $(\bar{\mu}, \hat{v})$ debate equilibria surviving the refinement is precisely the downward sloping thick line, that is, the upper boundary of the unrefined set.

Consider any $(p', q') \in (p, q)[\mathbf{1}]$ for which $(\bar{\mu}, \hat{v})$ is a bias-driven debate equilibrium (the information structure $(0.68, 0.30864)$ works for the refined, non-generic, case). Assume the realized profile of signals is $\mathbf{s} = (-1, 0, 0)$ so the right committee decision is $y$. Assume that the two uninformed individuals, $i = 2, 3$, are both $x$-biased. Then under the no-debate voting equilibrium $\mathbf{1}$, the committee unanimously votes for the right alternative $y$. Under the debate equilibrium $(\bar{\mu}, \hat{v})$, however, the realized debate is $\mathbf{m} = \mathbf{s} = (-1, 0, 0)$ but subsequent equilibrium voting has both individuals 2 and 3 voting for $x$, thus vetoing $y$ and leading to the wrong committee decision. $\square$

The reason for the error in the example establishing Proposition 5 is not hard to see. In the relevant information structure, the probability of any individual being informed is sufficiently low that a single noisy speech for $y$ is insufficient to offset any private bias for $x$. When there is no debate, however, the uninformed $x$-biased individuals condition on being pivotal, that is, on the event that both of the other committee members are surely voting for $y$, in which event there is positive probability on both individuals observing signals for $Y$ being the true state. On balance, the *ex ante* possibility of there being two signals in favour of $Y$ conditional on being pivotal without debate, is stronger support for choosing $y$ than knowing as a result of debate that there is at most one signal in favour of $Y$.

Proposition 5 implies that an analogous claim to Proposition 3 (which holds with or without the trembles refinement) is not available. The result does not imply that deliberation is on balance detrimental to the quality of committee decision making under unanimity rule and it seems unlikely that this is the case. What is true, is that, in comparison with majority rule, requiring unanimous voting induces quite distinct sorts of deliberation and incentives to share information in debate. And on balance, majority rule offers more opportunity for credible deliberation and symmetric information sharing.

# 6 Discussion

Despite the fact that the role of deliberation in agenda-setting *per se* may likely prove the most important, there is still a great deal to be learned about deliberation over fixed agendas. Assuming a fixed agenda, the particular issue we address in this paper concerns the connection between the voting rule adopted by a committee for making a decision and the character of any deliberation preceding the vote. Overall, the results point to majority rule being superior both with respect to the expected quality of committee decisions and to the quality of debate it induces. Specifically (with respect to pure strategy equilibria):

(1) for many circumstances, majority rule can result in full information equivalent debate equilibria, but there are no circumstances for which this is true of unanimity rule;

(2) with respect to making errors in common interest, debate weakly dominates no debate under majority rule but not under unanimity, where debate can result in errors that are not made without debate;

and

(3) with respect to making errors in bias alone, the only circumstances under which debate does not weakly dominate no debate under majority rule are those in which the separating debate equilibrium exists but is not played. And although there exist circumstances under which debate weakly improves on no-debate under unanimity rule, we do not yet know whether this is a general property of debate equilibria under unanimity.

The analysis underlying our results depends on what is, at least from a standard game-theoretic perspective, a fairly natural conception of committee deliberation, specifically, deliberation as strategic information transmission. And within this framework, there are some fairly obvious extensions, including sequential speechmaking, consequential variation in the relative weights individuals' place on private interests, and so on. However, the usual apparatus of incomplete information games may in fact to be too restrictive to address some of the important questions considered in the normative political theory literature. And a key issue in this regard concerns whether or not all consequential deliberation is inherently informational. If it turns out that in fact arguments predicated on strategic information transmission models fail to capture the salient features of committee deliberation precisely because these features are not intrinsically informational, then the relevance of our discussion to the normative literature becomes moot.

There seem to be two principal ways in which deliberation might not be informational. Loosely speaking, the first involves equilibrium selection in coordination games (Farrell, 1987; Rabin, 1994; Calvert and Johnson, 1998) and the second involves argument through analogy and precedent (Aragones, Gilboa, Postlewaite and Schmeidler, 2001).

Although it is surely the case that coordination and argument through analogy do not concern information of the sort considered in the model here, they are both intrinsically concerned with some form of informational imperfection. This is most evident for coordination games; here, no new information regarding the state of the world is produced in debate but the extent to which speech is informative is the extent to which any *strategic* uncertainty is resolved. Thus speech can lead to *ex post* Pareto efficiency gains by facilitating coordination on a particular equilibrium and, in the typical case where the distribution of payoffs is not neutral across equilibria, any tension in deliberation involves the equilibrium on which to coordinate.

Aragones, Gilboa, Postlewaite and Schmeidler (AGPS) observe that not all persuasive arguments involve changes in beliefs through information sharing. Rather, many arguments are by analogy, whereby the speaker makes explicit to the listener relations between known facts that the listener may not have seen. As an example, they suggest an individual, initially predisposed against US intervention in the Iraqi invasion of Kuwait, may be induced to change her mind after an analogy is drawn between Hussein's actions toward Kuwait and Hitler's actions toward Poland. It is, AGPS claim, perfectly reasonable to assume that while both individuals are fully aware of the cases involved, only one of them has made any connection between the two.

There is a strong intuition for analogies being important for debate and it seems apparent that the setting is not one usefully captured by orthodox Bayesian theory. Nevertheless, analogic arguments still seem to be fundamentally concerned with information transmission, albeit of qualitatively different sort to that in the standard framework: the speaker in the example is pointing out a connection of which the listener was previously unaware. From this perspective, information asymmetries remain critical to any notion of consequential debate and what AGPS, along with those looking at the role of debate in coordination games, make explicit is that we are going to have look for new tools if we hope to model all of the relevant forms such information asymmetries might take. On the other hand, if AGPS are correct in claiming both that information is not the issue and that it is the relations between known sets of facts, or "cases", that form the basis of much persuasive rhetoric, then models permitting failures of logical, as well as informational, omniscience are going to prove important. For it seems that logically omniscient individuals under complete and full information are going to know all possible connections between facts.

# 7 Appendix A: proofs

We first derive some important threshold inequalities exploited in some of the formal arguments below.

Given a message strategy $\mu$ and debate $\mathbf{m} \in \mathcal{M}^3_\mu$, any equilibrium vote strategy $\upsilon$ has to satisfy the pivotal voting constraints: that is, conditional on being pivotal at $\upsilon$, a $b$-biased agent $i$ who observes a signal $s \in \{-1, 0, 1\}$ weakly prefers to vote for $z$ rather than $z'$ under majority rule if and only if

$$E[U(z; b)|s, \mathbf{m}, \mu, z, \upsilon_{-i}, \mathbf{votepiv}] \geq E[U(z'; b)|s, \mathbf{m}, \mu, z', \upsilon_{-i}, \mathbf{votepiv}]$$

and, by definition of being pivotal, if the individual votes $z$ in this event then $z$ surely wins. With this in mind, let $b = z = y$, $z' = x$ and substitute for preferences $U(\cdot; y)$ into the inequality to yield

$$\begin{aligned} & E[U(y; y)|\cdot, \mathbf{votepiv}] - E[U(x; y)|\cdot, \mathbf{votepiv}] \\ = \; & \Pr[Y|s, \mathbf{m}, \mathbf{votepiv}] + \Pr[X|s, \mathbf{m}, \mathbf{votepiv}]\lambda \\ & - \Pr[X|s, \mathbf{m}, \mathbf{votepiv}](1 - \lambda) \\ = \; & \Pr[Y|s, \mathbf{m}, \mathbf{votepiv}] + (1 - \Pr[Y|s, \mathbf{m}, \mathbf{votepiv}])(2\lambda - 1), \end{aligned}$$

where the strategy pair $(\mu, \upsilon_{-i})$ is understood and, in obvious notation, we write $\Pr[Z|\cdot] \equiv \Pr[\omega = Z|\cdot]$, $Z \in \{X, Y\}$.[21] It follows that a $y$-biased individual votes for $y$ rather than for $x$ at $\upsilon$ only if

$$\lambda \geq \frac{1}{2}\left(1 - \frac{\Pr[Y|s, \mathbf{m}, \mathbf{votepiv}]}{1 - \Pr[Y|s, \mathbf{m}, \mathbf{votepiv}]}\right).$$

By Bayes rule,

$$\begin{aligned} & \Pr[Y|s, \mathbf{m}, \mathbf{votepiv}] \\ = \; & \frac{\Pr[Y|s]\Pr[\mathbf{votepiv}|Y, \mathbf{m}]}{\Pr[Y|s]\Pr[\mathbf{votepiv}|Y, \mathbf{m}] + \Pr[X|s]\Pr[\mathbf{votepiv}|X, \mathbf{m}]} \\ = \; & \frac{\Omega(s)}{\Omega(s) + \Phi(\mathbf{m})} \end{aligned}$$

where

$$\Omega(s) \equiv \frac{\Pr[Y|s]}{\Pr[X|s]} \text{ and } \Phi(\mathbf{m}) \equiv \frac{\Pr[\mathbf{votepiv}|X, \mathbf{m}]}{\Pr[\mathbf{votepiv}|Y, \mathbf{m}]}.$$

[21] An analogous inequality can be derived for the pivotal signaling constraints in similar fashion (although it is important in this case to fix the vote strategy $\upsilon$ across all individuals, including the one to whom a particular constraint applies).

So a $y$-biased individual votes for $y$ rather than for $x$ at $\upsilon$ only if

$$\lambda \geq \frac{1}{2}\left(1 - \frac{\Omega(s)}{\Phi(\mathbf{m})}\right).$$

Similarly, an $x$-biased ($b = x$) individual who observes signal $s$ weakly prefers to vote for $x$ at $\upsilon$ only if

$$\lambda \geq \frac{1}{2}\left(1 - \frac{\Phi(\mathbf{m})}{\Omega(s)}\right).$$

Further, if voting strategies are symmetric and committee decision making is by majority rule, the following are easily checked:

$$\begin{array}{ll} i. & \Phi(m, m', m'') = \Phi(m, m'', m'). \\ ii. & \Phi(m, m', m'') = \frac{1}{\Phi(-m, -m', -m'')} \\ iii & \quad\quad \Phi(0, m, -m) = 1 \end{array}$$

**Proof of Proposition 1** Let $\upsilon^0$ be any symmetric pure strategy voting equilibrium and, without loss of generality, let $i \in N$ have $y$-bias ($b_i = y$) and signal $s_i \in \{-1, 0, 1\}$. For $j \neq i$ and $k \neq i$, $(s_j, s_k)$ must satisfy exactly one of the following: (a) $s_j = s_k = 0$; (b) $s_j = -s_k \neq 0$; (c) $s_j = s_k \neq 0$; (d) $s_j + s_k = -1$; (e) $s_j + s_k = 1$. Then for each $\omega \in \{X, Y\}$,

$$\Pr[a|\omega] = (1-q)^2, \;\; \Pr[b|\omega] = q^2 p(1-p), \;\; \Pr[c|\omega] = \frac{1}{2}q^2(2p-1);$$

and

$$\Pr[d|X] = \Pr[e|Y] = \frac{1}{2}(1-q)q(1-p), \;\; \Pr[d|Y] = \Pr[e|X] = \frac{1}{2}(1-q)qp.$$

Now suppose $i$ is vote pivotal. Then $j \neq i$ and $k \neq i$ must be voting for different alternatives. Furthermore, $\upsilon^0$ symmetric implies that, conditional on $i$ being pivotal, (d) can be true of $(s_j, s_k)$ if and only if (e) can be true of $(s_j, s_k)$. Hence, although not every possibility in $\{$(a),...,(e)$\}$ need have strictly positive probability conditional on $i$ vote pivotal, $\upsilon^0$ symmetric implies

$$\Pr[\mathbf{votepiv}|\upsilon^0, Y] = \Pr[\mathbf{votepiv}|\upsilon^0, X].$$

By Bayes rule, therefore, $\Pr[Y|s_i, \upsilon^0, \mathbf{votepiv}]$ in this case is simply

$$\frac{\Pr[\mathbf{votepiv}|\upsilon^0, Y]\Pr[Y|s_i]}{\Pr[\mathbf{votepiv}|\upsilon^0]} = \Pr[Y|s_i].$$

Substituting for $U(\cdot; y)$ into the pivotal voting constraint (with debate ignored) and collecting terms, voting for $y$ is a best response for $i$ if and only if:

$$
\begin{aligned}
&E[U(y;y)|s, y, \upsilon^0_{-i}, \textbf{votepiv}] - E[U(x;y)|s, x, \upsilon^0_{-i}, \textbf{votepiv}] \\
&\qquad = \Pr[Y|s, \textbf{votepiv}] + \Pr[X|s, \textbf{votepiv}]\lambda - \Pr[X|s, \textbf{votepiv}](1-\lambda) \\
&\qquad = \Pr[Y|s, \textbf{votepiv}] + (1 - \Pr[Y|s, \textbf{votepiv}])(2\lambda - 1) \geq 0
\end{aligned}
$$

where the dependency on $\upsilon^0_{-i}$ is understood. It follows that a $y$-biased individual votes for $y$ rather than for $x$ at $\upsilon^0$ only if

$$
\begin{aligned}
\lambda &\geq \frac{1 - 2\Pr[Y|s_i, \textbf{votepiv}]}{2(1 - \Pr[Y|s_i, \textbf{votepiv}])} \\
&= \frac{(1 - 2\Pr[Y|s_i])}{2(1 - \Pr[Y|s_i])}.
\end{aligned}
$$

If $s_i = 1$, $\Pr[Y|s_i] = (1-p)$ and the constraint for voting $y$ is $\lambda \geq l_1(p)$; if $s_i \leq 0$, $\Pr[Y|s_i] \geq 1/2$ and the constraint for voting $y$ is $\lambda \geq 0$. This proves the proposition. $\square$

**Proof of Proposition 2** Let $(\mu, \upsilon)$ be a full information equivalent separating debate equilibrium at $(p, q)$. Given $\mu$ is separating in common interests, it is immediate from the definition of $l_1(p)$ that $\lambda \leq l_1(p)$ is necessary and sufficient for $\upsilon$ to satisfy the pivotal voting constraints and be full information equivalent voting. We therefore have to check the pivotal signaling constraints, given $\lambda \leq l_1(p)$.

Without loss of generality, consider a $y$-biased individual $i \in N$. It is straightforward to check that if $s_i = -1$ then $m_i = -1$ is the unique best response to $\mu_{-i}$. Suppose $i$ has signal $s_i = 0$. Given $(\mu_{-i}, \upsilon)$ and $s_i = 0$, it is clear that $i$ never strictly prefers sending message $m_i'' = 1$ rather than sending $m_i = 0$; and $i$ is willing to send the message $m_i = s_i = 0$ rather than deviate to a speech $m_i' = -1 < s_i$ if and only if

$$
E[U(z;y)|0, 0, \boldsymbol{\mu}_{-i}, \upsilon, \textbf{sigpiv}] \geq E[U(z';y)|0, -1, \boldsymbol{\mu}_{-i}, \upsilon, \textbf{sigpiv}].
$$

Given $(\mu, \upsilon)$, $i$ is signal pivotal at $s_i = 0$ between $m_i = 0$ and $m_i' = -1$ if either (a) both $j$ and $k$ are uninformed, have a bias for $x$, and send messages $m_j = m_k = 0$, or (b) both $j$ and $k$ are informed, have a bias for $x$, and send messages $m_j = -m_k = 1$, or (c) $j$ is uninformed and sends $m_j = s_j = 0$, $k$ is informed and sends message $m_k = s_k = 1$, and both $j, k$ have a bias for $y$. Suppose $i$ sends the truthful message $m_i = s_i = 0$. Then the committee decision is surely $x$. On the other hand, if $i$ sends the message $m_i' = -1$, the committee decision is surely $y$. With these remarks in mind, compute

$$
\Pr[Y|s_i, \mu_{-i}, \upsilon, \textbf{sigpiv}] =
$$

$$
\frac{\Pr[\textbf{sigpiv}|\boldsymbol{\mu}_{-i}, \upsilon, Y]\Pr[Y|s_i]}{\Pr[\textbf{sigpiv}|\mu_{-i}, \upsilon, Y]\Pr[Y|s_i] + \Pr[\textbf{sigpiv}|\mu_{-i}, \upsilon, X]\Pr[X|s_i]}
$$

where

$$\Pr[\mathbf{sigpiv}|\mu_{-i},\upsilon,Y] \equiv [\frac{1}{4}(1-q)^2+\frac{1}{2}q^2p(1-p)+\frac{1}{2}q(1-p)(1-q)],$$
$$\Pr[\mathbf{sigpiv}|\mu_{-i},\upsilon,X] \equiv [\frac{1}{4}(1-q)^2+\frac{1}{2}q^2(1-p)p+\frac{1}{2}qp(1-q)].$$

Since $\Pr[Y|s_i=0]=1/2$, $i$ is willing to send $m_i=0$ rather than $m_i'=-1$ only if

$$\begin{aligned}\lambda &\leq \frac{1-2\Pr[Y|0,\mu_{-i},\upsilon,\mathbf{sigpiv}]}{2(1-\Pr[Y|0,\mu_{-i},\upsilon,\mathbf{sigpiv}])} \\ &= \frac{q(1-q)(2p-1)}{[(1-q)^2+2q^2(1-p)p+2qp(1-q)]} \\ &< l_1(p).\end{aligned}$$

Now suppose, $s_i=1$. If ever $i$ prefers to send a message $m_i''=0$ rather than the message $m_i=1$, then $i$ surely prefers to send a message $m_i'=-1$ rather than the message $m_i=1$. So it suffices to identify when sending $m_i=1$ is a best response for $i$. Given $(\mu,\upsilon)$, $i$ is signal pivotal between $m_i=1$ and $m_i'=-1$ at events (a') both $j$ and $k$ are uninformed and send messages $m_j=m_k=0$, or (b') both $j$ and $k$ are informed and send messages $m_j=-m_k=1$, or (c') $j$ is uninformed and sends $m_j=s_j=0$, $k$ is informed and sends message $m_k=s_k=1$, and both $j,k$ have a bias for $y$, or (d') where $j$ is uninformed and sends $m_j=s_j=0$, $k$ is informed and sends message $m_k=s_k=-1$, and both $j,k$ have a bias for $x$. Then whichever event obtains, if $i$ sends the truthful message $m_i=s_i=1$, the committee decision is surely $x$ and, if $i$ sends the message $m_i'=-1$, the committee decision is surely $y$. Thus

$$\Pr[\mathbf{sigpiv}|\mu_{-i},\upsilon,Y] \equiv$$
$$[(1-q)^2+\frac{1}{2}q^2p(1-p)+\frac{1}{4}q(1-p)(1-q)+\frac{1}{4}qp(1-q)],$$

and

$$\Pr[\mathbf{sigpiv}|\mu_{-i},\upsilon,X] \equiv$$
$$[(1-q)^2+\frac{1}{2}q^2(1-p)p+\frac{1}{4}qp(1-q)+\frac{1}{4}q(1-p)(1-q)].$$

Rehearsing the same argument as before, *mutatis mutandis*, yields that $i$ is willing to send $m_i=1$ rather than $m_i'=-1$ only if

$$\begin{aligned}\lambda &\leq \frac{1-2\Pr[Y|1,\mu_{-i},\upsilon,\mathbf{sigpiv}]}{2(1-\Pr[Y|1,\mu_{-i},\upsilon,\mathbf{sigpiv}])} \\ &= \frac{(2p-1)}{2p}=l_1(p).\end{aligned}$$

Therefore the binding signal pivot constraint is that on the uninformed individual, in which case there exists a full information equivalent separating debate equilibrium if and only if

$$\lambda \leq \frac{q(1-q)(2p-1)}{[(1-q)^2 + 2q^2(1-p)p + 2qp(1-q)]}.$$

Maximizing the RHS of this inequality with respect to $q$ and $p$ in turn, substituting back and taking limits appropriately yields Proposition 2(1) and 2(2), completing the proof. □

**Proof of Lemma 1** Let $(\mu, \upsilon)$ be an SP debate equilibrium and suppose the lemma is false at $(\mu, \upsilon)$. Assume individual $i$ is uninformed ($s_i = 0$), has bias $b_i = y$ and that $(m_i, m_j, m_k) = (-1-1, -1)$. Given $\mu$ is semi-pooling in common interests, a message $m = -1$ is sent in debate only if the sender has a signal $s = -1$ or is $y$-biased and uninformed. By supposition

$$\upsilon(y, 0, -1, -1, -1) = x.$$

By $\mu$ semi-pooling, it must be that for $j, k \neq i$,

$$(s_j, s_k) \in \{(0,0), (0,-1), (-1,-1)\}.$$

If ever $s_j = 0$, then $\mu$ semi-pooling implies $j$ is $y$-biased and the supposition requires $j$ to vote surely for $x$. Hence, individual $i$ cannot be vote pivotal if $(s_j, s_k) = (0,0)$. And if $i$ is vote pivotal under majority rule and $(s_j, s_k) = (0,-1)$, it must be that $k$ votes for $y$; and if $(s_j, s_k) = (-1,-1)$, $j, k$ must (given majority rule) have opposite bias. In any case, the pivotal voting constraints imply that a $y$-biased individual is willing to vote for $x$ at $\mathbf{m} = (-1,-1,-1)$ if and only if

$$\lambda \leq \frac{1}{2}\left(1 - \frac{\Omega(0)}{\Phi(-1,-1,-1)}\right)$$

where, for any signal $s$ and debate $\mathbf{m} \in \mathcal{M}_\mu^3$,

$$\Omega(s) \equiv \frac{\Pr[Y|s]}{\Pr[X|s]} \text{ and } \Phi(\mathbf{m}) \equiv \frac{\Pr[\mathbf{votepiv}|X, \mu, \upsilon, \mathbf{m}]}{\Pr[\mathbf{votepiv}|Y, \mu, \upsilon, \mathbf{m}]}.$$

Given $\Omega(0) = 1$, there exist $\lambda \in (0, 1/2)$ satisfying the inequality only if $\Phi(-1,-1,-1) > 1$. Because $i$ can be pivotal at $\mathbf{m} = (-1,-1,-1)$ given $s_i = 0$ only at the events identified above, we have

$$\begin{aligned}\Phi(-1,-1,-1) &= \frac{\left[\frac{1}{2}(1-q)q(1-p) + \frac{1}{2}q^2(1-p)^2\right]}{\left[\frac{1}{2}(1-q)qp + \frac{1}{2}q^2p^2\right]} \\ &= \frac{(1-p)(1-qp)}{p(1-q+qp)}.\end{aligned}$$

But $p \geq 1/2$ implies $\Phi(-1,-1,-1) < 1$. This contradiction proves the lemma. $\square$

**Proof of Lemma 2** The pivotal voting constraints imply a $y$-biased individual is willing to vote for $y$ at $\mathbf{m} \in \mathcal{M}_\mu^3$ given a signal $s$ if and only if

$$\lambda \geq \frac{1}{2}\left(1 - \frac{\Omega(s)}{\Phi(\mathbf{m})}\right).$$

Similarly, an $x$-biased individual who observes signal $s$ weakly prefers to vote for $x$ at $\mathbf{m} \in \mathcal{M}_\mu^3$ if and only if

$$\lambda \geq \frac{1}{2}\left(1 - \frac{\Phi(\mathbf{m})}{\Omega(s)}\right).$$

By assumption, $p > 1/2$; hence, $\Omega$ is strictly decreasing in $s$. The claims now follow directly. $\square$

**Proof** Suppose $(m_i, m_j, m_k) = (1,-1,-1)$ and assume individual $i$ who sends message $m_i = 1$ cannot be pivotal and that both agents $j$ and $k$ always vote for $x$ irrespective of their bias and signal. Then it must be the case that

$$v(y,-1,-1,1,-1) = x. \tag{$*$}$$

Consider such a $y$-biased individual who has observed signal $s = -1$ and sent message $m = -1$ and observes a split debate $(1,-1)$ and who is supposed to vote for $x$. There can be no event such that this agent's vote is pivotal for this debate since otherwise he must vote for y. To see this note that the observed split debate and the assumption of the SP signalling strategy implies at most one other agent has observed the signal 1 so it follows that if there is a positive probability the agent is pivotal he should vote for y. To ensure such an agent votes for $x$ it must be the case that his vote cannot be pivotal. But then, since the other agent sending message $m_k = -1$ is always voting for $x$ by assumption we get the requirement that

$$v(1,1,1,-1,-1) = x.$$

Symmetry and anonymity implies

$$v(-1,-1,-1,1,1) = y. \tag{$**$}$$

But equations (*) and (**) imply a violation of debate monotonicity. $\square$

**Proof of Proposition 3** Fix any feasible information structure $(p, q)$. By Proposition 1, there is a unique equilibrium in pure strategies without debate: when $\lambda < l_1(p)$, all informed individuals surely vote their signal and all uninformed individuals vote their bias; when $\lambda > l_1(p)$ all individuals vote their bias. Let $v^0$ denote this no-debate voting strategy and let $(\mu, \upsilon)$ be any pure strategy debate equilibrium (in undominated strategies and subject to the maintained technical refinement). Then the proposition is trivial if $\mu$ is either separating or pooling in common interest. Suppose $(\mu, \upsilon)$ is a semi-pooling debate equilibrium.

Under a semi-pooling equilibrium, all individuals offer make speeches that reveal how they would have voted without debate. For a committee decision distinct to the no-debate decision, therefore, at least one person must change their vote as a consequence of the debate. As a consequence of debate, that is, either an informed individual votes against her signal or an uninformed individual votes against her bias. Moreover, if the outcome is going to be worse with debate than without, it must be that an individual who changes her vote switches to the worse outcome. Let $y$ be the right outcome; then the committee can make an error in common interest by choosing $x$ following debate only if $y$ is defined by unanimous induced preferences at $\mathbf{s}$. So there can be an error in common interests only if (up to permutations)

$$\mathbf{s} \in \{(-1,-1,-1),(0,-1,-1),(1,-1,-1),(-1,0,0)\}.$$

We consider each case in turn. Throughout, the SP debate equilibrium $(\mu, \upsilon)$ is fixed and taken as understood.

(I) $(s_1, s_2, s_3) = (-1,-1,-1)$. Under $\upsilon^0$ all individuals vote for $y$ and, given the signal profile and definition of $\mu$, the debate must be $\mathbf{m} = (-1,-1,-1)$. Consequent on $\mathbf{m}$, therefore, there are essentially two possible voting outcomes $\mathbf{v} = (v_1, v_2, v_3)$ that result in a mistake:

(a) $\mathbf{v} = (x, x, x)$. In this case all agents are supposed to vote for $x$. By Lemma 1 $\upsilon(y, 0, -1, -1, -1) = y$ and, therefore, by signal monotonicity (Lemma 2.1), $\upsilon(y, -1, -1, -1, -1) = y$. It follows that $b_i = x$ for all $i$, so we must have $\upsilon(x, -1, -1, -1, -1) = x$. Consider any $x$-biased agent who is supposed to vote for $x$ here. For this debate, there is a positive probability of being pivotal and $\Phi(-1,-1,-1)$ is defined. Specifically,

$$\begin{aligned}
\Phi(-1,-1,-1) &= \frac{\Pr[\mathbf{votepiv}|X, \mu, \upsilon, -1, -1, -1]}{\Pr[\mathbf{votepiv}|Y, \mu, \upsilon, -1, -1, -1]} \\
&= \frac{\frac{1}{2}q(1-q)(1-p) + \frac{1}{2}q^2(1-p)^2}{\frac{1}{2}q(1-q)p + \frac{1}{2}q^2p^2} \\
&= (1-p)\,\frac{1-qp}{p\,(1-q+qp)}.
\end{aligned}$$

Now $\Omega(-1) = p/\,(1-p)$ so the agent is willing to vote for $x$ only if

$$\begin{aligned}
\lambda &\geq \frac{1}{2}\left(1 - \frac{(1-p)}{p}\frac{(1-p)(1-qp)}{p\,(1-q+qp)}\right) \\
&= \frac{1}{2}\frac{(2p-1)(1-qp\,(1-p))}{p^2\,(1-q+qp)} \\
&> 1
\end{aligned}$$

But then, by Proposition 1, the no-debate equilibrium $v^0$ requires all individuals to vote their bias which makes $x$ the right outcome and contradicts the supposition of an error here.

(b) $\mathbf{v} = (x, x, y)$ or $\mathbf{v} = (y, x, x)$. For either of these possibilities to constitute equilibrium behaviour here requires $v(x, 0, -1, -1, -1) = x$. But then the same logic as for (a) applies and we obtain a contradiction.

(II) $(s_1, s_2, s_3) = (0, -1, -1)$. It must be the case that the uninformed agent is $x$-biased since otherwise all the messages are $-1$ and the argument in case I(a) applies. As indicated (and without loss of generality), assume $s_1 = 0$ and therefore, by $\mu$ semi-pooling, $(m_1, m_2, m_3) = (1, -1, -1)$. By Lemma 3, if 1 is not pivotal then the right decision must be made. So if there is an error, 1 must have positive probability of being pivotal here. And for 1's vote to be pivotal it must be the case that individuals 2 and 3 are different (if they have the same bias, send the same message and observe the same messages from others then they vote the same way). But since $m_2 = m_3 = -1$, $\mu$ semi-pooling implies $s_j \leq 0$, $j = 2, 3$, and moreover $m_j = -1$ and $s_j = 0$ imply $b_j = y$. There can be only one such agent $j \in \{2, 3\}$ in the pair if a vote is pivotal, so the other agent, $k$, must have observed $s_k = -1$. In this case the uninformed $x$-biased agent, $i = 1$, who sends message $m_1 = 1$ must believe that, conditional on being pivotal, exactly one other agent $k$ has observed signal $s_k = -1$. In which case, by $l_1 > \lambda$, individual $i = 1$ prefers to vote for $y$.

(III) $(s_1, s_2, s_3) = (1, -1, -1)$. Then $(m_1, m_2, m_3) = (1, -1, -1)$. By the same argument as for (II), if a vote is pivotal it must be the case that agents 2 and 3 are different. Consequently, at least one of these agents must have observed the signal $s_j = -1$; let $j = 2$. It follows that agent 1 cannot be $y$-biased: for if $b_1 = y$, then he would vote for $y$ conditional on being pivotal because he knows the third agent has not seen $s = 1$. Either $b_2 = x$ or $b_3 = x$; assume $b_2 = x$. Since 2 and 3 must be different, it must be that $b_3 = y$ and $s_3 \leq 0$. It follows that $k = 3$ votes for $y$, implying both individuals 1 and 2 are voting for $x$. Now individual 1 is $x$-biased and $s_1 = 1$; therefore 1 prefers to vote for $x$ only if

$$\lambda \geq \frac{1}{2}\left(1 - \frac{\Phi(1, -1, -1)}{\Omega(1)}\right).$$

Because $\lambda < l_1$ the above inequality can be satisfied only if $\Phi(1, -1, -1) > 1$ but, given the voting strategies described above,

$$\Phi(1, -1, -1) = \frac{p\left(q^2(1-p)^2 + \frac{1}{2}q(1-q)(1-p)\right)}{(1-p)\left(q^2p^2 + \frac{1}{2}q(1-q)p\right)} < 1$$

since $p > 1/2$.

(IV) $(s_1, s_2, s_3) = (0, 0, -1)$. By $\mu$ semi-pooling, if $b_1 = b_2 = y$ then $m_1 = m_2 = -1$ and, therefore, by Lemmas 1 and 2(1), both individuals surely vote $y$. On the other hand, because informed individuals vote their signal and uninformed individual vote their bias when $\lambda < l_1$ and there is no debate, if $b_1 = b_2 = x$ then the decision under no debate is $x$ and evidently a debate equilibrium cannot do worse. To obtain a mistake therefore, it is necessary that $b_1 \neq b_2$; without loss of generality, assume $b_1 = x$ and $b_2 = y$. Then the debate is $(m_1, m_2, m_3) = (1, -1, -1)$. By Lemma 3, if there is an error there must be positive probability of $i = 1$ being pivotal at this debate. But then individuals 2 and 3 must be voting differently and therefore, by $m_2 = m_3$ and $M_{-2} = M_{-3}$, have different biases. By semi-pooling debate, $m_j = -1$ implies either $s_j = 0$ and $b_j = y$ or $s_j = -1$. Hence, individual 1 knows surely that $s_2 + s_3 \leq -1$ in which case, since $\lambda < l_1$ and $s_1 = 0$, 1 surely votes $y$.

Because (I) through (IV) exhaust the possibilities for errors in common interest, we are done. □

The following lemma is useful for proving Corollary 1. Let $\mu$ be the semipooling message strategy.

**Lemma 4** *If $(\mu, \upsilon)$ and $(\mu, \upsilon')$ are both symmetric and debate monotonic semipooling debate equilibria under which uninformed individuals vote against their bias on hearing a split debate. Then, along the equilibrium path, $\upsilon = \upsilon'$ and equilibrium voting decisions are described by the profile R of Table 1 in the text.*

**Proof**. Consider equilibrium path voting behaviour. By hypothesis, along the equilibrium path uninformed individuals vote against their bias on hearing a split debate $(-1, 1)$; that is,

$$\upsilon(y, 0, -1, -1, 1) = x \text{ and } \upsilon(x, 0, 1, -1, 1) = y \tag{1}$$

By (1) and debate monotonicity,

$$\upsilon(y, 0, -1, 1, 1) = x \text{ and } \upsilon(x, 0, 1, -1, -1) = y \tag{2}$$

By (1) and Lemma 2 (signal monotonicity),

$$\upsilon(y, 1, 1, -1, 1) = x \text{ and } \upsilon(x, -1, -1, -1, 1) = y \tag{3}$$

By (3) and debate monotonicity,

$$\upsilon(y, 1, 1, 1, 1) = x \text{ and } \upsilon(x, -1, -1, -1, -1) = y \tag{4}$$

By (3) and Lemma 2 (bias monotonicity),

$$\upsilon(x, 1, 1, -1, 1) = x \text{ and } \upsilon(y, -1, -1, -1, 1) = y \tag{5}$$

Similarly, by (4) and Lemma 2 (bias monotonicity),

$$\upsilon(x, 1, 1, 1, 1) = x \text{ and } \upsilon(y, -1, -1, -1, -1) = y \tag{6}$$

And by Lemma 1,

$$\upsilon(y, 0, -1, -1, -1) = y \text{ and } \upsilon(x, 0, 1, 1, 1) = x \tag{7}$$

There remain two (equilibrium path) decisions to be determined; specifically, for each $z \in \{x, y\}$

$$\upsilon(z, -1, -1, 1, 1) \text{ and } \upsilon(z, 1, 1, -1, -1)$$

Suppose first that individual $i \in N$ has $\upsilon(x, -1, -1, 1, 1) = y$. Then both of the other two committee members observe a split debate. Hence, (1) through (7) imply there exists a unique event at which $i$'s vote is pivotal: there exists an uninformed ($s_j = 0$) $x$-biased individual $j$ who has sent message $m_j = 1$, hears a split debate and votes for $y$; and there exists an informed ($s_k = 1$) individual $k$ who has sent message $m_k = 1$, hears a split debate and votes for $x$. But then $i$'s unique undominated vote decision is to vote for $x$. Therefore,

$$\upsilon(x, -1, -1, 1, 1) = x$$

in which case, by symmetry

$$\upsilon(y, 1, 1, -1, -1) = y$$

Now suppose that individual $i \in N$ has $\upsilon(y, -1, -1, 1, 1) = x$. Then $i$'s vote is pivotal in exactly the same case as above; but since $i$ is now presumed $y$-biased, we conclude

$$\upsilon(y, -1, -1, 1, 1) = y$$

so by symmetry

$$\upsilon(x, 1, 1, -1, -1) = x.$$

And because there exist no further unspecified equilibrium path voting decisions, this proves the lemma. □

**Proof of Corollary 1** To prove the result, it suffices to show there exists a symmetric and debate monotonic semipooling debate equilibrium at $(\lambda, p, q)$ in which uninformed individuals vote against their bias on hearing a split debate only if there exists a separating debate equilibrium at $(\lambda, p, q)$. A necessary condition for any such semipooling debate equilibrium to exist is for the pivotal constraints to hold along equilibrium path. So consider a $y$-biased individual who has signal $s = 0$, sends message $m = -1$ and observes a split debate $(-1, 1)$. By hypothesis, $\upsilon(y, 0, -1, -1, 1) = x$. By Lemma 4, the unique equilibrium voting path in any such semipooling debate equilibrium is described by the strategy R in Table 1. Therefore, there are three events at which the vote of an uninformed individual $i$, having sent message $m_i = -1$ and observed a split debate $(m_j, m_k) = (-1, 1)$, is pivotal:

*Either* both $j$ and $k$ are uninformed: $j$ is $y$-biased, $M_{-j} = 0$ and votes $x$; $k$ is $x$-biased, $M_{-k} = -2$ and votes $y$;

*Or* $j$ is uninformed, $y$-biased and votes $x$ given $M_{-j} = 0$; $k$ is informed with $s_k = 1$, $y$-biased and votes $y$ given $M_{-k} = -2$;

*Or* $j$ is informed with $s_j = -1$, $y$-biased and votes $y$ given $M_{-j} = 0$; $k$ is informed with $s_k = 1$, $x$-biased and votes $x$ given $M_{-k} = -2$.

Substituting into the pivotal voting constraint and collecting terms, we find $v(y, 0, -1, -1, 1) = x$ is an undominated best response only if

$$\lambda \leq \lambda_R \equiv \frac{q(1-q)(2p-1)}{2[(1-q)^2 + qp(1-q) + 2q^2 p(1-p)]}.$$

From the proof to Proposition 2, the binding pivotal signaling constraint for the separating debate equilibrium requires

$$\lambda \leq \lambda_S \equiv \frac{q(1-q)(2p-1)}{[(1-q)^2 + 2qp(1-q) + 2q^2 p(1-p)]}.$$

Hence, at any information structure $(p, q) \in (\frac{1}{2}, 1) \times (0, 1)$,

$$\lambda_R < \lambda_S \Leftrightarrow 0 < (1-q)^2 + 2q^2 p(1-p)$$

which is obviously true. This fact proves the result. $\square$

**Proof of Proposition 4** Suppose by way of contradiction that $(\mu, v)$ is a separating debate equilibrium. Then no new information is revealed by the fact that a vote is pivotal and, therefore, the sincere voting strategy is weakly dominant; in particular, given $\mathbf{m} \in \mathcal{M}_\mu^3$, $\mathbf{m} = \mathbf{s}$ and $\lambda < l_1$ implies

$$v(y, s_i, \mathbf{m}) = \begin{cases} y \text{ if } s_i + M_{-i} \leq 0 \\ x \text{ otherwise} \end{cases} \quad \text{and } v(x, s_i, \mathbf{m}) = \begin{cases} y \text{ if } s_i + M_{-i} < 0 \\ x \text{ otherwise} \end{cases}$$

with the sincere strategy being defined analogously for $\lambda > l_1$. We show that an uninformed ($s = 0$) individual with $y$-bias strictly prefers to send message $m = -1$ to message $m = 0$, thus violating the relevant pivotal signaling constraint for $\mu$ separating in common interests. Because the event that an individual is signal or vote pivotal under unanimity implies that both the other committee members are making similar decisions, to prove the result it suffices to check the case $\lambda < l_1$. Given $\lambda < l_1$, $\mu$ separating and $v$ sincere, an uninformed $y$-biased individual $i$ is signal pivotal between $m_i = 0$ and $m_i' = -1$ under unanimity rule with status quo $x$ if and only if (a) $(m_j, m_k) = (s_j, s_k) = (0, 0)$ and at least one of $j, k$ has an $x$-bias, or (b) $(m_j, m_k) = (s_j, s_k) = (-1, 1)$ and at least one of $j, k$ has an $x$-bias. Therefore,

$\Pr[Y|s_i, \mu_{-i}, \upsilon, \mathbf{sigpiv}] =$

$$\frac{\Pr[\mathbf{sigpiv}|\boldsymbol{\mu}_{-i}, \upsilon, Y] \Pr[Y|s_i]}{\Pr[\mathbf{sigpiv}|\mu_{-i}, \upsilon, Y] \Pr[Y|s_i] + \Pr[\mathbf{sigpiv}|\mu_{-i}, \upsilon, X] \Pr[X|s_i]}$$

where

$$\begin{aligned}
\Pr[\mathbf{sigpiv}|\mu_{-i}, \upsilon, Y] &\equiv [\frac{3}{4}(1-q)^2 + \frac{3}{2}q^2 p(1-p)], \\
\Pr[\mathbf{sigpiv}|\mu_{-i}, \upsilon, X] &\equiv [\frac{3}{4}(1-q)^2 + \frac{3}{2}q^2 (1-p)p].
\end{aligned}$$

By $\upsilon$ sincere, in either event (a) or (b), individual $i$ votes her bias whatever message she delivers. On the other hand, both $j$ and $k$ vote surely for $y$ in these events if $m_i' = -1$ and at least one of them votes for $x$ otherwise. Therefore, since $\Pr[Y|s_i = 0] = 1/2$, substituting into the relevant signal pivot constraint implies that $i$ is willing to send $m_i = 0$ rather than $m_i' = -1$ only if

$$\lambda \leq \frac{1 - 2\Pr[Y|0, \mu_{-i}, \upsilon, \mathbf{sigpiv}]}{2(1 - \Pr[Y|0, \mu_{-i}, \upsilon, \mathbf{sigpiv}])} = 0$$

which contradicts $\lambda > 0$. $\square$

# 8 Appendix B: refinement and derivations

In this appendix we define the technical (trembles) equilibrium refinement and describe the approach to identifying particular classes of equilibria discussed in the paper. With some abuse to the notation in the text, it is useful to begin by redefining some variables. Fix an individual $i \in N$ and hereafter suppress any individual-specific subscripts. Assume also that committee decision making is by majority rule; similar constructions apply to the case of unanimity.

Let $b = -1$ if the individual's bias is for $y$ and let $b = 1$ if her bias is for $x$. Similarly, let $\omega = -1$ if the state of the world is $Y$ and let $\omega = 1$ if the state is $X$. Let $m, m'$ etc denote the relevant individual's messsage in any debate and let $\theta, \rho$ denote the messages of the other two committee members; by convention, when writing any debate $\mathbf{m} = (m, \theta, \rho) \in \mathcal{M}^3 = \{-1, 0, 1\}^3$, the relevant individual's message is always listed first.

## 8.1 Refinement

For any profile $(b, s, m, \theta, \rho) \in \{-1, 1\} \times \{-1, 0, 1\}^4$ and any $z \in \{x, y\}$, let

$$\upsilon(b, s, m, \theta, \rho) \in [0, 1]$$

be the probability that an individual with bias $b$ and signal $s$, having sent debate message $m$ and heard messages $\theta, \rho$, votes for alternative $y$. Since there is no abstention, $1-\upsilon(b, s, m, \theta, \rho)$ is the probability that the individual votes for $x$. Similarly, for any $(b, s) \in \{-1, 1\} \times \{-1, 0, 1\}$ and any $m \in \{-1, 0, 1\}$ let

$$\mu(b, s, m) \in [0, 1]$$

be the probability that an individual with bias $b$ and signal $s$ sends debate message $m$; by assumption, $\sum_{m \in \{-1,0,1\}} \mu(b, s, m) = 1$.

With this notation, an anonymous message strategy is a triple

$$\mu = (\mu(b, s, -1), \mu(b, s, 0), \mu(b, s, 1))$$

and an anonymous voting strategy is simply a pair of vote-probabilities sufficiently described by

$$\upsilon = \upsilon(b, s, m, \theta, \rho).$$

Let $(\mu, \upsilon)$ be an equilibrium in pure strategies, i.e. the adding up constraints are satisfied and $\mu(b, s, m) \in \{0, 1\}$ for each message $m$ and $\upsilon(b, s, m, \theta, \rho) \in \{0, 1\}$. It is irrelevant to apply any refinement to separating debate equilibria as there is no out-of-equilibrium behaviour to worry about. So assume for this discussion that $\mu$ is semi-pooling in common interests. Then the only messages supposed to be sent in equilibrium are $m = -1$ and $m' = 1$. However, if ever a message $m = 0$ is observed in a semi-pooling equilibrium, we assume all individuals surely identify the (out of equilibrium) message with the message $m = -1$. Now consider the voting strategy $\upsilon$.

Given $\upsilon$, define a perturbed voting strategy component-wise by

$$\upsilon(b, s, m, \theta, \rho; \varepsilon) = \left\{ \begin{array}{ll} 1-\varepsilon & \text{if } \upsilon(b, s, m, \theta, \rho) = 1 \\ \varepsilon & \text{otherwise} \end{array} \right. ,$$

where $\varepsilon > 0$ and small. For each $\varepsilon$, let

$$\upsilon(\varepsilon)) = ((\upsilon(b, s, m, \theta, \rho; \varepsilon))$$

Then the pure strategy pair $(\mu, \upsilon)$ survives the technical refinement (individual-invariant trembles) if

$$\lim_{\varepsilon \to 0} (\mu, \upsilon(\varepsilon)) = (\mu, \upsilon).$$

## 8.2 Derivations

For each rule and any conjectured equilibrium strategy pair $(\mu, \upsilon)$, we have to identify the signal pivot and vote pivot constraints for each possible event. Typically, there are a great many such events to check To see why, consider an individual with bias $b$ and signal $s$ who is supposed to send message $m$ in debate; then there are two possible deviations from $m$ and, for each deviation, there are multiple distinct pivot events. And given a realized debate, the vote pivot constraints for the individual have to be checked for each possible message he might have sent, both in and out of equilibrium, and for each possible debate that might be realized. Finally, this family of constraints has to be checked for consistency. Not surprisingly, the algebra becomes very cumbersome and tedious very rapidly. We therefore wrote a program using the Maple V symbolic manipulation package in Scientific Workplace 4.1 to identify the relevant pivot events and do the algebra. This is available from the authors on request.

# 9 References

Aragones, E., I. Gilboa, A. Postlewaite and D. Schmeidler. 2001. Rhetoric and Analogies. U.Penn WP (May).

Austen-Smith, D. 1990a. Information transmission in debate. *American Journal of Political Science, 34:124-52.*

Austen-Smith, D. 1990b. Credible debate equilibria. *Social Choice and Welfare,* 7:75-93.

Austen-Smith, D. 1993a. Interested experts and policy advice: multiple referrals under open rule. *Games and Economic Behavior, 5:3-43.*

Austen-Smith, D. 1993b. Information acquisition and orthogonal argument. In W. Barnett, M. Hinich and N. Schofield (eds) *Political Economy: Institutions, Competition and Representation* Cambridge:CUP.

Bohman, J. and W. Rehg (eds) 1997. *Deliberative Democracy: Essays on Reason and Politics* Cambridge MA:MIT Press.

Calvert, R. and J. Johnson. 1998. Rational argument, political argument and democratic deliberation. Paper presented to the 1998 Annual Meetings of the APSA.

Christiano, T. 1997. The significance of public deliberation. In J. Bohman and W. Rehg (eds) *Deliberative Democracy* Cambridge MA:MIT Press.

Cohen, J. 1989. Deliberation and Democratic Legitimacy. In, A.Hamlin and P.Pettit (eds), *The Good Polity*, Oxford:Basil Blackwell.

Diermeier, D. and T. Feddersen. 2000. Information and congressional hearings. *American Journal of Political Science* 44:

Doraszelski, U., D. Gerardi and F. Squintani. 2001. Communication and voting with double-sided information. Northwestern University WP (April).

Duggan, J. and C. Martinelli 2001. A Bayesian model of voting in juries. *Games and Economic Behavior* 37:259-94.

Elster, J. (ed) 2000. *Deliberative Democracy* Cambridge:CUP.

Elster, J. 1997. The market and the forum: three varieties of political theory. In J. Bohman and W. Rehg (eds) *Deliberative Democracy* Cambridge MA:MIT Press.

Estlund, D 1997. Beyond fairness and deliberation: the epistemic dimension of democratic authority. In J. Bohman and W. Rehg (eds) *Deliberative Democracy* Cambridge MA:MIT Press.

Farrell, J. 1987. Cheap talk, coordination and entry. *RAND Journal of Economics* 18:34-39.

Farrell, J. 1983. Meaning and credibility in cheap talk games. *Games and Economic Behavior, 5:514-31.*

Feddersen, T. and W. Pesendorfer. 1996. The swing voter's curse. *American Economic Review* 86:408-24.

Feddersen, T. and W. Pesendorfer. 1997. Voting behavior and information aggregation in elections with private information. *Econometrica, 65:1029-58.*

Gauss, G. 1997. Reason, justification and consensus: why democracy can't have it all. In J. Bohman and W. Rehg (eds) *Deliberative Democracy* Cambridge MA:MIT Press.

Glazer, J. and A. Rubinstein. 2001. Debates and decisions: on a rationale of argumentation rules *Games and Economic Behavior* 36:158-73.

Gutmann, A. and D. Thompson. 1995. Moral disagreement in a democracy. *Social Philosophy and Policy, 12:87-110.*

Karni, E. and Z. Safra. 2002. Individual sense of justice: a utility representation. *Econometrica, 70:263-84.*

Knight, J. and J. Johnson. 1997. What sort of equality does deliberative democracy require? In J. Bohman and W. Rehg (eds) *Deliberative Democracy* Cambridge MA:MIT Press.

Ladha, K. 1992. The Condorcet Jury Theorem, free speech and correlated votes. *American Journal of Political Science, 36*:6*17-34.*

Lipman, B. and D. Seppi. 1995. Robust inference in communication games with partial provability. *Journal of Economic Theory*, 66:370-405.

Manin, B. 1987. On legitimacy and political deliberation. *Political Theory, 15:338-68.*

Matthews, S. and A. Postlewaite. 1992. On modeling cheap talk in Bayesian games. In, J.Ledyard (ed), *The Economics of Informational Decentralization: Complexity, Efficiency and Stability*, Amsterdam:Kluwer Nijhoff.

McLean, I. and F. Hewitt. 1994. *Condorcet: Foundations of Social Choice and Political Theory*. Aldershot:Edward Elgar Publishing.

McLennan, A. 1998. Consequences of the Condorcet Jury Theorem for beneficial information aggregation by rational players. *American Political Science Review*, *92*:4*13*-18.

Miller, D. 1993. Deliberative democracy and social choice. In, D.Held (ed), *Prospects for Democracy*, Stanford:Stanford University Press.

Ottaviani, M. and P. Sorensen. 2001. Information aggregation in debate: who should speak first? *Journal of Public Economics*, 81:393-422.

Rabin, M. 1994. A model of pregame communication. *Journal of Economic Theory*, 63:370-91.

Sanders, L. 1997. Against deliberation. *Political Theory*, 25:347-76.

Spector, D. 2000. Rational debate and one-dimensional conflict. *Quarterly Journal of Economics*, 467:181-200.

Sunstein, C. 1993. Democracy and shifting preferences. In, D.Copp, J.Hampton and J.Roemer, *The Idea of Democracy*, Cambridge:CUP.