Deliberation, Unanimity Rule and Majority Rule [1]

David Austen-Smith

and

Tim Feddersen

MEDS, Kellogg School of Management
Northwestern University
Evanston, IL 60208

First version, December 2002
This revision, August 2004

Abstract

A deliberative committee is a group of at least two individuals who first debate about what alternative to choose prior to these same individuals voting to determine the choice. We demonstrate in a very general setting that the condition under which unanimity can support full information revelation in debate amounts to it being common knowledge that all committee members invariably share identical preferences over alternatives. It follows that if ever there exists an equilibrium with fully revealing dedebate under unanimity rule, there exists an equilibrium with fully revealing debate under any voting rule. Moreover, the converse is not true.

---

[1] Earlier versions of this paper were titled "The Inferiority of Deliberation under Unanimity Rule".

1 Introduction

A very widely held intuition in the literature on jury and committee deliberation is that requiring a unanimous vote provides the strongest incentives for those involved to share their opinions and decision-relevant information most fully. Particularly direct statements of this intuition include

> "It should be remembered that veto power or unanimity represents a constraint that induces deliberation: when parties can block outcomes, actors have incentives to find reasons that are convincing to all, not just to the majority" (Eriksen, 2001:15/16)

and

> "The necessity of a consensus of all jurors which flows from the requirement of unanimity, promotes deliberation and provides some insurance that the opinions of each of the jurors will be heard and discussed" (South Australian High Court, 1993; quoted in Walker and Lane, 1994:2)

Similar perspectives on the connection between unanimity and deliberation also color much of the more normative literature on deliberative democracy (see, for example, Shapiro, 2002; Dryzek and List, 2002; and the essays in Bohman and Rehg, 1997 and in Elster, 2000). It is, however, a flawed intuition. Although those who support the alternative requiring unanimous consent have a strong incentive to reveal information supporting that alternative, they likewise have a correspondingly strong incentive not to reveal information favoring the *status quo*. And the latter incentive in turn generates an externality, rendering information suspect. We show that the only circumstances under which unanimous voting induces full information sharing in a deliberative setting is when there can be no conflict about which collective choice is the right one at every possible state. But given such preference homogeneity, deliberation can be fully informative under any voting rule at all. Moreover, other rules can support full information sharing even when the unanimity rule cannot.

The existing formal literature concerned with strategic deliberation and multi-person committee voting under incomplete information is as yet small.[2] Austen-Smith (1990a,b) considers the role of debate in a spatial model of endogenous agenda-setting under majority rule; Calvert and Johnson (1998) look at the coordinating role of debate in a complete information model of committee decision-making; Meirowitz (2004) also worries about debate and coordination but under incomplete information; and Hafer and Landa (2003) develop a non-Bayesian model of argument.[3] The most closely related contributions to the current paper, however, are Doraszelski, Gerardi and Squintani (2001), Austen-Smith and Feddersen (2004) and Coughlan (2000). All three of these papers address, *inter alia*, the relationship between the unanimity voting rule and deliberation; connecting the salient results to our argument is deferred for now. Finally, Gerardi and Yariv (2004) adopt a quite different approach, either to the papers cited here or to this paper. In particular, unlike the focus on full information revelation in what follows, Gerardi and Yariv do not consider any qualitative properties of deliberation *per se*. Framing the issue as one of mechanism design under incomplete information, they instead study the abstract relationships between sets of sequential equilibrium outcomes achievable through unmediated cheap-talk communication in voting games, proving in particular that the set of sequential equilibrium outcomes achievable by augmenting a voting game under a q-rule with a prior communication stage is constant in the quota q, for all q-rules other than

---

[2] There is a complementary formal literature concerned with n-person debate aimed at influencing an uninformed monopolistic decision-maker. Examples include Glazer and Rubinstein (2001), Ottaviani and Sorensen (2001), Lipman and Seppi (1995), Diermeier and Feddersen (2001) and Austen-Smith (1993a, 1993b). Key differences between the papers cited in the text and those falling within this complementary class are that, in the latter, the set of individuals deliberating does not coincide with the set of individuals responsible for making a decision and there is no explicit concern with strategic voting.

[3] Although neither paper is concerned with strategic debate, Aragones, Gilboa, Postlewaite and Schmeidler (2001) and Gilboa and Vieille (2004), like Hafer and Landa, explore non-Bayesian models of argument. Aragones *et al* consider various formalizations of the role of analogy and Gilboa and Vieille ask what, if any, restrictions does the case-based decision model of (sincere) individual votes place on the majority preference relation.

unanimity.[4] A key feature of their argument is the fact that all individuals *always* voting unanimously is consistent with sequential equilibrium under any non-unanimous rule. But in the case that deliberation prior to a vote results in complete information among committee members at the time of the vote, prescribing that individuals surely vote the same way in any decision amounts to prescribing that individuals use dominated voting strategies.

The next section describes the formal model and main result for unanimity rule: in the absence of fully homogenous preferences, voting under unanimity rule is incapable of supporting complete information sharing in debate. The subsequent section shows by example that there are non-unanimous rules for which this is not the case. A final section briefly connects our results to the formal literature and concludes.

2 Model and result

Consider a committee $N = \{1, 2, ..., n\}$, $n \geq 2$, that has to choose an alternative $z \in \{x, y\}$; let x be the *status quo* policy. Each individual $i \in N$ has private information $(b_i, s_i) \in B \times S$, where b_i is a preference parameter, or *bias*, and s_i is a *signal* regarding the alternatives. Assume the sets B and S are finite and common across individuals $i \in N$. Write $B^n \equiv \mathbf{B}$ and $S^n \equiv \mathbf{S}$; a *situation* is any pair $(\mathbf{b}, \mathbf{s}) \in \mathbf{B} \times \mathbf{S}$, where $\mathbf{b} = (b_1, ..., b_n)$, $\mathbf{s} = (s_1, ..., s_n)$. And with a convenient abuse of language, any profile $\mathbf{s} \in \mathbf{S}$ is a *state*. Let $p(\mathbf{b}, \mathbf{s})$ be the probability that situation $(\mathbf{b}, \mathbf{s}) \in \mathbf{B} \times \mathbf{S}$ obtains..

For any committee member $i \in N$, i's preferences over $\{x, y\}$ depend exclusively on i's own bias $b_i \in B$ and on the state $\mathbf{s} \in \mathbf{S}$: given a bias b and state $\mathbf{s}$, an individual's payoff from a committee decision $z \in \{x, y\}$ is written $u(y, b, \mathbf{s})$. We assume there are no dogmatic or partisan types; that is, for any bias $b \in B$ there is a nonempty subset of states $\mathbf{S}_b \subset \mathbf{S}$ such that $\mathbf{s} \in \mathbf{S}_b$ implies $u(y, b, \mathbf{s}) > u(x, b, \mathbf{s})$ and $\mathbf{s} \notin \mathbf{S}_b$ implies $u(y, b, \mathbf{s}) < u(x, b, \mathbf{s})$. To avoid

---

[4] A *q-rule* is a voting rule such that if at least $q \geq 1$ committee members vote for y against x, then y is the committee decision.

trivialities we assume that every situation occurs with positive probability and, because the concern here is with unanimity rule, that there always exist states at which all members prefer alternative y.

Axiom 1 (Full Support) *For all* $(\mathbf{b}, \mathbf{s}) \in \mathbf{B} \times \mathbf{S}$*,* $p(\mathbf{b}, \mathbf{s}) > 0$*.*

Axiom 2 (Consensus) *For all* $\mathbf{b} = (b_1, ..., b_n) \in \mathbf{B}$*,* $\mathbf{S}(\mathbf{b}) \equiv \cap_{i \in N} \mathbf{S}_{b_i} \neq \emptyset$*.*

Given that the committee is to choose from a fixed binary agenda, it is fairly natural to interpret signals $s \in S$ as constituting more or less evidence for choosing one or other of the two altenatives. Consequently, assume that the set of signals S is ordered by a binary relation, $\succ$, such that the following monotonicity condition obtains.

Axiom 3 (Monotonicity) *For any* $s, s' \in S$ *such that* $s \succ s'$ *and* $\mathbf{s}_- \in S^{n-1}$*, let* $\mathbf{s} = (\mathbf{s}_-, s) \in \mathbf{S}$ *and* $\mathbf{s}' = (\mathbf{s}_-, s') \in \mathbf{S}$*. Then* $u(y, b, \mathbf{s}) > u(y, b, \mathbf{s}')$ *and* $u(x, b, \mathbf{s}) < u(x, b, \mathbf{s}')$ *for any* $b \in B$*.*

In words, suppose there is a pair of states that differ only in that some member has observed $s \in S$ in the first state and $s' \in S$; then $s \succ s'$ implies that s is stronger information than s' in favor of y and against x. And notice that this axiom also builds in a degree of symmetry: any individual's relative evaluation of the two alternatives is monotone in signals whatever the individual's bias b and irrespective of exactly which committee member receives what signal.

The committee chooses an outcome by voting under unanimity rule. That is, x is the outcome unless every member of the committee votes for y. Prior to voting we assume there is a deliberation phase in which every member of the committee can simultaneously send a message m to every other member of the committee. For any $i \in N$, bias $b \in B$ and signal $s \in S$, let M be the set of available messages where M is an arbitrary, uncountably infinite set. A *message strategy* for $i \in N$ is a function, $\mu_i : B \times S \to M$. A message profile $\mathbf{m} = (m_1, m_2, ..., m_n) \in M^n \equiv \mathbf{M}$ is a *debate.*

Definition 1 *A message strategy profile* $\boldsymbol{\mu}$ *is* fully revealing *if, for all* $i \in N$*, for all pairs of distinct signals* $s, s' \in S$*,* $[\cup_{b \in B} \mu_i(b, s)] \cap [\cup_{b \in B} \mu_i(b, s')] = \emptyset$*.*

As defined here, fully revealing message strategies may or may not reveal information about individual biases. Because individuals' preferences depend only on the state and on their own bias, if a debate fully reveals the state then additional information about others' biases is decision-irrelevant. Thus the key feature of a fully revealing message strategy is that it provides full information about the speaker's signal.

Consistent with the motivation for the paper, our focus is on deliberation that yields all individuals' information being shared prior to the voting stage; that is, on fully revealing debates. In this context, there is no loss of generality in associating messages directly with the information they are presumed to report, so assume $B \times S \subset M$. Then $\boldsymbol{\mu}$ is fully revealing if, for all $i \in N$ and all $(b, s) \in B \times S$, $\mu_i(b, s) = s$.

Definition 2 *A committee is* minimally diverse *if and only if there exist* $b, b' \in B$ *such that* $\mathbf{S}_b \neq \mathbf{S}_{b'}$*.*

In words, a committee is minimally diverse if its membership exhibits preference heterogeneity at least to the extent that there is some pair of individual bias parameters that disagree about the states in which alternative y should be selected. Under the full support assumption, it is possible for all individuals to exhibit the same bias and, therefore, the only committees that are not minimally diverse are committees in which there is *never* any disagreement about when alternative y is the best choice ($\mathbf{S}_b = \mathbf{S}_{b'}$ for all $b, b' \in B$).

For all $\mathbf{b} \in \mathbf{B}$, let $\mathbf{T}^0(\mathbf{b}) \equiv \mathbf{S}(\mathbf{b})$ and, for any $k = 1, 2, \ldots$, recursively define the sets

$$\mathbf{T}^k(\mathbf{b}) = \{\mathbf{s} \notin \cup_{l=1}^{l=k} \mathbf{T}^{k-l}(\mathbf{b}) | \exists s, s' \in S : s' \succ s,\ (\mathbf{s}_-, s) = \mathbf{s},\ (\mathbf{s}_-, s') = \mathbf{s}' \text{ and } \mathbf{s}' \in \mathbf{T}^{k-1}(\mathbf{b})\}.$$

Thus $\mathbf{T}^1(\mathbf{b})$ is the set of states not in $\mathbf{T}^0(\mathbf{b})$ such that, given the realized bias profile $\mathbf{b}$, changing any one person's information from s to s' results in a state in $\mathbf{T}^0(\mathbf{b}) \equiv$; $\mathbf{T}^2(\mathbf{b})$ is the set of states not in $\mathbf{T}^1(\mathbf{b})$ such that changing any one person's information from s to s' results in a state in ; and so on. Informally, the set $\mathbf{T}^k(\mathbf{b})$ is the set of states such that there is a path of k single coordinate changes of information that lead to a state at which y is preferred unanimously. Since S and N are finite it follows that

$$\cup_{k=0,1,\ldots,n}\mathbf{T}^k(\mathbf{b}) = \mathbf{S}.$$

For example, suppose $n = 3$, $S = \{0,1\}^3$ and $\mathbf{S}(\mathbf{b}) = \{(1,1,1)\}$. Then

$$\begin{aligned} T^0(\mathbf{b}) &= \{(1,1,1)\} \\ \mathbf{T}^1(\mathbf{b}) &= \{(0,1,1),(1,0,1),(1,1,0)\} \\ \mathbf{T}^2(\mathbf{b}) &= \{(1,0,0),(0,1,0),(0,0,1)\} \\ \mathbf{T}^3(\mathbf{b}) &= \{(0,0,0)\}. \end{aligned}$$

The following property of minimally diverse committees in environments satisfying the three axioms is useful for proving the main theorem.

Lemma 1 *Assume full support, consensus and monotonicity. In a minimally diverse committee there exists a bias profile* $\mathbf{b} = (\mathbf{b}_-, b, b') \in \mathbf{B}$ *and a state* $\mathbf{s} \in \mathbf{T}^1(\mathbf{b})$ *such that* $\mathbf{s} \notin \mathbf{S}_b$ *but* $\mathbf{s} \in \mathbf{S}_{b'}$.

Proof Let $\mathbf{b} = (\mathbf{b}_-, b, b') \in \mathbf{B}$ (where, by an abuse of notation, it is understood that $\mathbf{b}_- \in B^{n-2}$); by consensus, $\mathbf{S}(\mathbf{b}) \neq \emptyset$. First assume there is a state $\mathbf{s} \in \mathbf{S}_b \cap \mathbf{T}^{k+1}(\mathbf{b})$. By full support and definition of $\mathbf{T}^k(\mathbf{b})$, there exists a signal $s' \succ s$ such that $(\mathbf{s}_-, s') = \mathbf{s}' \in \mathbf{T}^k(\mathbf{b})$; moreover, by monotonicity, $\mathbf{s}' \in \mathbf{S}_b$. Hence, $\mathbf{s} \in \mathbf{S}_b \cap \mathbf{T}^{k+1}(\mathbf{b})$ implies there exists a state $\mathbf{s}' \in \mathbf{S}_b \cap \mathbf{T}^k(\mathbf{b})$. Now suppose b is such that, for any $\mathbf{s} \in \mathbf{T}^k(\mathbf{b})$, $\mathbf{s} \notin \mathbf{S}_b$. Then by the previous argument, there can be no $\mathbf{s} \in \mathbf{T}^{k+1}(\mathbf{b})$ such that $\mathbf{s} \in \mathbf{S}_b$. Hence, $\mathbf{S}_b \cap \mathbf{T}^1(\mathbf{b}) = \emptyset$ implies $\mathbf{S}_b \cap \mathbf{T}^k(\mathbf{b}) = \emptyset$ for all $k > 1$ in which case, because $\cup_{k=0,1,\ldots,n}\mathbf{T}^k(\mathbf{b}) = \mathbf{S}$, it must be that $\mathbf{S}_b = \mathbf{S}(\mathbf{b})$. It follows that if, contrary to the lemma, for all $\mathbf{b} \in \mathbf{B}$ there exists no $\mathbf{s} \in \mathbf{T}^1(\mathbf{b})$ and components b, b'

of $\mathbf{b}$ such that $\mathbf{s} \notin \mathbf{S}_b$ but $\mathbf{s} \in \mathbf{S}_{b'}$, then $\mathbf{S}_b = \mathbf{S}(\mathbf{b})$ for all components of $\mathbf{b}$, violating minimal diversity. $\square$

A *voting strategy* for member $i \in N$ is a function $\nu_i : B \times S \times \mathbf{M} \rightarrow \{x, y\}$ that maps every debate into a voting decision. A *fully revealing debate equilibrium* is a Perfect Bayesian Equilibrium $(\boldsymbol{\mu}, \boldsymbol{v}) = ((\mu_1, \ldots, \mu_n), (\nu_1, \ldots, \nu_n))$ such that $\boldsymbol{\mu}$ is fully revealing and $\boldsymbol{v}$ is a profile of weakly undominated voting strategies.

Theorem 1 *Assume full support, consensus and monotonicity. There exists a fully revealing debate equilibrium if and only if the committee is not minimally diverse.*

Proof (Necessity) In any fully revealing debate equilibrium, the restriction to weakly undominated voting strategies implies $\nu_i(b, s, \mathbf{m}) = y$ if and only if $(\mathbf{s}_{-i}, s) \in \mathbf{S}_b$, where $\mathbf{s}_{-i} = \mathbf{m}_{-i}$ for every $i \in N$ and $b \in B$. It follows that a member's voting strategy does not depend on the message she sends in debate. Consider the deliberation stage and, by way of contradiction, suppose $\boldsymbol{\mu}$ is fully revealing yet the committee is minimally diverse. Then, given the behavior at the voting stage, fully revealing message strategies constitute an equilibrium if and only if, for every $i \in N$ and every $(b_i, s_i) \in B \times S$, it is the case that

$$EU(m_i = s_i, b_i, s_i) - EU_i(m_i = s', b_i, s_i) \geq 0 \text{ for any } s' \in M \backslash \{s_i\} \tag{1}$$

where $EU(m_i, b_i, s_i) =$

$$\sum_{\mathbf{b}_{-i} \in B^{n-1}} \sum_{\mathbf{s}_{-i} \in S^{n-1}} p(\mathbf{b}_{-i}, \mathbf{s}_{-i} | b_i, s_i) \left[\Pr(x | \mathbf{b}, \mathbf{s}, m_i) u(x, b_i, \mathbf{s}) + \Pr(y | \mathbf{b}, \mathbf{s}, m_i) u(y, b_i, \mathbf{s}) \right]$$

and $\Pr(z | \mathbf{b}, \mathbf{s}, m_i)$ is the probability that $z \in \{x, y\}$ is the committee decision given bias profile $\mathbf{b} = (\mathbf{b}_{-i}, b_i)$, state $\mathbf{s} = (\mathbf{s}_{-i}, s_i)$ and debate $(\mathbf{m}_{-i}, m_i) = (\mathbf{s}_{-i}, m_i)$. Fix $i \in N$ and let $(b_i, s_i) = (b, s)$; for any $s' \in M \backslash \{s\}$, define the function

$$\varphi_{(b,s)}(s, s'; \mathbf{b}_{-i}, \mathbf{s}_{-i}) \equiv \left[\Pr(x | \mathbf{b}, \mathbf{s}, s) - \Pr(x | \mathbf{b}, \mathbf{s}, s') \right] \left[u(x, b, \mathbf{s}) - u(y, b, \mathbf{s}) \right]$$

with $\mathbf{b} = (\mathbf{b}_{-i}, b)$ and $\mathbf{s} = (\mathbf{s}_{-i}, s)$. Then we can rewrite (1) equivalently as requiring that for all $(b, s) \in B \times S$ and all $s' \in M \backslash \{s\}$,

$$\sum_{\mathbf{b}_{-i} \in B^{n-1}} \sum_{\mathbf{s}_{-i} \in S^{n-1}} p(\mathbf{b}_{-i}, \mathbf{s}_{-i} | b, s) \varphi_{(b,s)}(s, s'; \mathbf{b}_{-i}, \mathbf{s}_{-i}) \geq 0. \tag{2}$$

By assumption, $\boldsymbol{\mu}_{-i}$ is fully revealing of all others' signals and, by the preceding argument on ν_i, for all messages $m_i \in M$ and all bias profiles $(\mathbf{b}_{-i}, b) \in \mathbf{B}$, $(\mathbf{s}_{-i}, s) \in \mathbf{S} \backslash \mathbf{S}_b$ implies $\Pr(x | (\mathbf{b}_{-i}, b), (\mathbf{s}_{-i}, s), m_i) = 1$. Similarly, for any state $(\mathbf{s}_{-i}, s) \in \mathbf{S} \backslash (\mathbf{S}(\mathbf{b}) \cup \mathbf{T}^1(\mathbf{b}))$ it must be that $\Pr(x | (\mathbf{b}_{-i}, b), (\mathbf{s}_{-i}, s), m_i) = 1$. Given $(b_i, s_i) = (b, s)$, therefore, for all $s' \in M \backslash \{s\}$ and all $\mathbf{b}_{-i} \in B^{n-1}$,

$$(\mathbf{s}_{-i}, s) \in \mathbf{S} \backslash \left[\mathbf{S}(\mathbf{b}) \cup \mathbf{T}^1(\mathbf{b}) \cup \mathbf{S}_b\right] \Rightarrow \varphi_{(b,s)}(s, s'; \mathbf{b}_{-i}, \mathbf{s}_{-i}) = 0. \tag{3}$$

The preceding argument implies that an individual i with bias b can change the outcome by switching from message s to some $s' \neq s$ only in situations $(\mathbf{b}, \mathbf{s})$ such that $(\mathbf{s}_{-i}, s) \in \mathbf{S}(\mathbf{b}) \cup \left(\mathbf{T}^1(\mathbf{b}) \cap \mathbf{S}_b\right)$. For all $\mathbf{b} \in \mathbf{B}$, define

$$X_i(\mathbf{b}, s, s') = \{(\mathbf{s}_{-i}, s) \in \mathbf{S}(\mathbf{b}) | (\mathbf{s}_{-i}, s') \notin \mathbf{S}(\mathbf{b})\}$$

to be the set of states such that if an individual i who is supposed to report s instead reports s' then, conditional on $\mathbf{b}$, the outcome changes from y to x. Similarly, define

$$Y_i(\mathbf{b}, s, s') = \{(\mathbf{s}_{-i}, s) \in \left(\mathbf{T}^1(\mathbf{b}) \cap \mathbf{S}_b\right) | (\mathbf{s}_{-i}, s') \in \mathbf{S}(\mathbf{b})\}$$

to be the set of states in which i prefers y and, if i is supposed to report s but instead reports s' at $\mathbf{b}$, the outcome changes from x to y. Note that, by monotonicity, if $Y_i(\mathbf{b}, s, s') \neq \emptyset$ for some $\mathbf{b} \in \mathbf{B}$, then $X_i(\mathbf{b}, s, s') = \emptyset$ for all $\mathbf{b} \in \mathbf{B}$ and, if $X_i(\mathbf{b}, s, s') \neq \emptyset$ for some $\mathbf{b} \in \mathbf{B}$, then $Y_i(\mathbf{b}, s, s') = \emptyset$ for all $\mathbf{b} \in \mathbf{B}$. That is, $Y_i(\mathbf{b}, s, s') \neq \emptyset$ for some $\mathbf{b} \in \mathbf{B}$ implies that s' is stronger evidence for y than s, whereas $X_i(\mathbf{b}, s, s') \neq \emptyset$ for some $\mathbf{b} \in \mathbf{B}$ implies s' is weaker evidence for y than s. By monotonicity both statements cannot be true. For any $\mathbf{b} \in \mathbf{B}$ and $s, s' \in S$, let

$$Z_i^-(\mathbf{b}, s, s') \equiv \{\mathbf{s}_{-i} \in S^{n-1} | (\mathbf{s}_{-i}, s) \in \left[Y_i(\mathbf{b}, s, s') \cup X_i(\mathbf{b}, s, s')\right]\}.$$

Collecting terms and using (3), we can rewrite the incentive compatibility constraint (2) as requiring, for all $i \in N$, $(b, s) \in B \times S$ and $s' \in M \backslash \{s\}$,

$$\sum_{\mathbf{b}_{-i} \in \mathbf{B}^-} \sum_{\mathbf{s}_{-i} \in Z_i^-(\mathbf{b}, s, s')} p(\mathbf{b}_{-i}, \mathbf{s}_{-i}, | b, s) \varphi_{(b,s)}(s, s'; \mathbf{b}_{-i}, \mathbf{s}_{-i}) \geq 0. \tag{4}$$

By Lemma 1 and full support, minimal diversity implies there is a $(\mathbf{b}_{-i}, b) \in \mathbf{B}$ and a pair of signals $s, s' \in S$ such that $Y_i((\mathbf{b}_{-i}, b), s, s') \neq \emptyset$ and $X_i((\mathbf{b}_{-i}, b), s, s') = \emptyset$. By definition, $(\mathbf{s}_{-i}, s) \in Y_i((\mathbf{b}_{-i}, b), s, s')$ implies $u(x, b, (\mathbf{s}_{-i}, s)) < u(y, b, (\mathbf{s}_{-i}, s))$ and $\Pr(x | (\mathbf{b}_{-i}, b), \mathbf{s}, s) - \Pr(x | (\mathbf{b}_{-i}, b), \mathbf{s}, s') = 1$. Hence, for all $(\mathbf{b}_{-i}, b) \in \mathbf{B}$,

$$\mathbf{s}_{-i} \in Z_i^-(\mathbf{b}, s, s') \Rightarrow \varphi_{(b,s)}(s, s'; \mathbf{b}_{-i}, \mathbf{s}_{-i}) < 0.$$

But then the incentive compatibility conditions are surely violated, contradicting the existence of a fully revealing debate equilibrium in any minimally diverse committee. This proves necessity.

(Sufficiency) Assume the committee is not minimally diverse. Then for all $b \in B$ and all $\mathbf{b} = (\mathbf{b}_-, b) \in \mathbf{B}$, $\mathbf{S}_b = \mathbf{S}(\mathbf{b})$. In this case there is no $\mathbf{b} \in \mathbf{B}$ and pair of signals $s, s' \in S$ such that $Y_i(\mathbf{b}, s, s') \neq \emptyset$ for any $i \in N$. Since incentive compatibility is assured for any $i \in N$, $b \in B$ and pair of signals $s, s' \in S$ such that $X_i(\mathbf{b}, s, s') \neq \emptyset$ and $Y_i(\mathbf{b}, s, s') = \emptyset$, full revelation is an equilibrium strategy. This completes the proof. □

Thus the circumstances under which unanimity rule promotes fully revealing deliberation are confined to those in which it is common knowledge that the committee is homogenous with respect to preferences over alternatives. Moreover, it is easy to see that the theorem holds in the case that the true bias profile $\mathbf{b} \in \mathbf{B}$ is common knowledge.[5] We close this section by recording an easy implication of Theorem 1; although technically straightforward, the corollary is substantively consequential.

---

[5] This is essentially a matter of notation: fix a bias profile $\mathbf{b} = (b_1, \ldots, b_n)$, suppose $\mathbf{b}$ is common knowledge and let $B = \{b_1, \ldots, b_n\}$. Then the definitions and analysis go through on replacing references to "biases $b, b' \in B$" with references to "individuals $i, j \in N$ with biases $b_i, b_j \in B$", and so on.

Let $q \in \{1, 2, \ldots, n\}$ and recall that a *q-rule* is a voting rule such that if at least $q \geq 1$ committee members vote for y against x, then y is the committee decision. Unanimity rule is a q-rule with $q = n$. Then noting that the sufficiency argument for Theorem 1 does not depend in any substantive way on the use of unanimity rule, this argument can be applied directly to any q-rule to yield the following corollary.

Corollary 1 *Assume full support, consensus and monotonicity. If there exists a fully revealing debate equilibrium under unanimity rule then there exists a fully revealing debate equilibrium under all q-rules.*

In other words, because committees that are not minimally diverse unanimously agree on the preferred alternative in every possible situation, such committees can always support fully revealing deliberation whatever voting rule is used to finalize a decision.

We now show that Theorem 1 is not true for at least one q-rule with $q < n$.

3 Example: fully revealing deliberation under majority rule

Consider a three person committee, $N = \{1, 2, 3\}$ and a set of situations $\mathbf{B} \times \mathbf{S} = \{x, y\} \times \{-1, 0, 1\}$.[6] Given a state $\mathbf{s} \in \mathbf{S}$, an individual $i \in N$ with bias $b_i \in B$ has a payoff from a committee choice of $z \in \{x, y\}$ given by

$$u(z, b_i, \mathbf{s}) = \lambda U(z, b_i) + (1 - \lambda) \Pr(z^* | \mathbf{s}),$$

where $\lambda \in (0, 1/2)$, $U(z, b_i) = 1$ if $z = b_i$, and $U(z, b_i) = 0$ otherwise. The remaining term, $\Pr(z^*|\mathbf{s})$, is the probability that z is the "socially correct" or "fair" available choice (e.g. maximizes aggregate utility or minimizes some measure of inequality) given state $\mathbf{s}$. The intended interpretation of these preferences is that the individual has both private and public interests. Thus, λ is the the weight an individual places on his or her own private interests in

---

[6] The example and proposition below are adapted from Austen-Smith and Feddersen (2004).

the committee decision, whereas $(1-\lambda)$ is the weight he or she places on making the socially correct outcome. The restriction of λ to the interval $(0, 1/2)$ insures that both considerations can be decision-relevant to the individual: $\lambda = 0$ means the individual has no effective private interest and $\lambda \geq 1/2$ means that any interest the individual has in the public well-being are invariably swamped by his or her private interests.

Of course, assuming all three individuals place the same relative weight on their private interest and share a common criterion of what constitutes socially correct is not terribly plausible; indeed, much interesting debate is precisely about these issues. However, since the aim here is to establish the claim that Theorem 1 does not extend to (in particular) majority rule, including such complications is unnecessary.

It remains to specify the information structure $p(\mathbf{b}, \mathbf{s})$ on $\mathbf{B} \times \mathbf{S}$. It is convenient (for the example) to assume that individual biases are drawn independently of individual signals, and that both biases and signals are iid across individuals. Specifically, for all individuals i, assume that

$$\Pr(b_i = x) = \Pr(b_i = y) = 1/2;$$

that the common prior belief about socially correct outcomes is uniform, so $\Pr[x^*] = 1/2$; and that

$$\begin{aligned}
\Pr(s_i &= 0|x^*) = \Pr(s_i = 0|y^*) = 1 - f \\
\Pr(s_i &= 1|x^*) = \Pr(s_i = -1|y^*) = fg \\
\Pr(s_i &= -1|x^*) = \Pr(s_i = 1|y^*) = f(1-g)
\end{aligned}$$

where $0 < f < 1$ and $1/2 < g < 1$. Then, for instance, we compute[7]

$$p((x, y, x), (0, 1, -1)) = \frac{1}{32}(1-f)f^2$$

---

[7] The probability that $(b_1, s_1) = (x, 0)$ is $\frac{1}{2}(1-f)$; the probability that $(b_2, s_2) = (y, 1)$ is $\frac{1}{2}f[\frac{1}{2}g + \frac{1}{2}(1-g)] = \frac{1}{4}f$ and, by symmetry, the probability that $(b_3, s_3) = (x, -1)$ is also $\frac{1}{4}f$. Applying independence then yields the reported probability.

The idea behind the signal structure is that with probability $1 - f$ an individual receives an uninformative signal, $s_i = 0$, about which outcome is fair and, with probability f, he or she receives an informative but noisy signal, $s_i \in \{-1, 1\}$, about what is fair. Each of these informative signals could occur whichever is the socially correct or fair choice in any state but, conditional on receiving an informative signal, the signal $s_i = 1$ (respectively, -1) is more likely to be received if the truly fair choice is x (respectively, y): that is, $1/2 < g < 1$. The signal structure is common knowledge but the realizations $(b_1 s_1)$, (b_2, s_2), (b_3, s_3) are private information to the relevant individuals.

By construction, the example $(N, \mathbf{B}, \mathbf{S}, p)$, satisfies *full support, consensus, monotonicity* and the committee is *minimally diverse* for all $\lambda > 0$. By Theorem 1, therefore, there exists no fully revealing debate equilibrium under unanimity. Nevertheless, we can prove that so long as individuals do not put too much weight on their private relative to their collective interests then majority rule does provide incentives for full information revelation in debate.

Proposition 1 *For any* $(f, g) \in (0, 1) \times (1/2, 1)$ *there is a unique value* $\lambda(f, g) \in (0, [2g-1]/2g)$ *such that there exists a fully revealing debate equilibrium under majority rule if and only if* $\lambda \leq \lambda(f, g)$.

A proof of this proposition is confined to the Appendix. Figure 1 (derived under the assumption that $\lambda = 1/10$) illustrates that the set of environments under which fully revealing debate equilibria exist can be quite extensive.

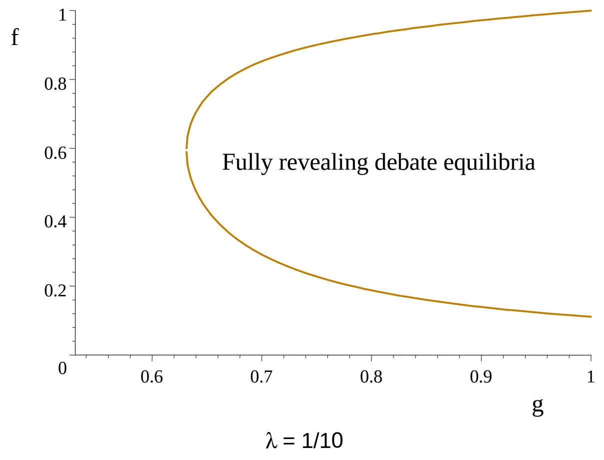

Figure 1: Fully revealing debate equilibria under majority rule

4 Conclusion

As remarked earlier, the most closely connected papers to the current contribution are Doraszelski, Gerardi and Squintani (2001), Austen-Smith and Feddersen (2004) and Coughlan (2000). Assuming cheap-talk debate, Doraszelski, Gerardi and Squintani (2001: Proposition 3) prove an impossibility result for fully revealing equilibria in a two-person model with deliberation over a fixed binary agenda with unanimity rule; Austen-Smith and Feddersen (2002: Proposition 4) establish a similar claim for a particular three-person committee; and Coughlan (2001: Proposition 5) finds sufficient conditions for a fully revealing debate among jurors. The no-separation in debate results in Doraszelski, Gerardi and Squintani (2001) and Austen-Smith and Feddersen (2004), and the separation result in Coughlan (2001) are immediate corollaries of Theorem 1: by construction, the committee is minimally diverse both in Doraszelski, Gerardi and Squintani and in Austen-Smith and Feddersen, and the sufficient conditions for Coughlan's claim insure the committee cannot be minimally diverse.

In contrast to a common wisdom, unanimous voting can, under very general conditions, create incentives for individuals to reveal less information in debate than they would be willing to reveal under a non-unanimous rule. Moreover, when circumstances are such that fully revealing deliberation is possible under unanimity, then it is likewise possible under all voting rules; the converse, however, is not true. Finally, while we examine a model with cheap talk the basic insight extends to a setting in which agents are sometimes constrained to tell the truth. As long as agents are capable of hiding some decision relevant information that supports the *status quo*, unanimity rule can provide the incentive to do so.

5 Appendix

The following preliminaries are useful for proving Proposition 1. First, let $z, z' \in \{x, y\}$ and consider an individual with bias b and signal s. Then substituting for payoffs we find

$$E[u(z,b,\mathbf{s})|s,\cdot] \geq E[u(z',b,\mathbf{s})|s,\cdot] \Leftrightarrow \lambda \leq \frac{2\Pr(z^*|s,\cdot)-1}{2\Pr(z^*|s,\cdot)-1+(U(z',b)-U(z,b))} \tag{5}$$

where $\Pr(z^*|s,\cdot)$ denotes the probability that alternative z is the socially correct choice given the individual's signal s and any additional information inferred from the behavior of others and so forth.

Now consider only symmetric equilibria, in that individuals' strategies depend on their respective biases and signals and not at all on their names; thus $\mu_i = \mu_j$ and $\upsilon_i = \upsilon_j$ for all $i, j \in N$. Then given a message strategy μ and debate $\mathbf{m} \in \mathcal{M}^3_\mu$, any equilibrium vote strategy υ has to satisfy the *pivotal voting constraints*: that is, conditional on being pivotal at υ, a b-biased agent i who observes a signal $s \in \{-1, 0, 1\}$ weakly prefers to vote for z rather than z' if and only if

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

where $\mathbf{vpiv}$ denotes the event of being pivotal at the voting stage and, by definition of being pivotal, if the individual votes z in this event then z surely wins.

Similarly, given a vote strategy υ, any equilibrium message strategy μ has to satisfy the *pivotal message constraints*: that is, conditional on being pivotal at μ, a b-biased agent i who observes a signal $s \in \{-1, 0, 1\}$ weakly prefers to send the message m rather than m' if and only if

$$E[u(z,b,\mathbf{s})|s,m,\mu_{-i},\upsilon,\mathbf{mpiv}(m,m')] \geq E[u(z',b,\mathbf{s})|s,m',\mu_{-i},\upsilon,\mathbf{mpiv}(m,m')]$$

where $\mathbf{mpiv}(m, m')$ denotes the event that the individual is message pivotal and messages m and m' produce different outcomes (respectively, z, z') at the voting stage.

Finally, fix $\mathbf{s}$ and let $S \equiv \sum_{i \in N} s_i$ be the sum of all individuals' signals and suppose S is known surely to an individual with bias b and signal s. Then it follows from the assumption of a uniform prior on the socially correct alternative and the inequality (5) that $\lambda < [2g-1]/2g$ implies all individuals strictly prefer x (respectively, y) when $S \geq 1$ (respectively, $S \leq -1$).

Proof of Proposition 1. Let (μ, υ) be a fully revealing debate equilibrium at the information structure $p = (f, g)$. Given μ is fully revealing, it is immediate that $\lambda \leq [2g-1]/2g$ is necessary and sufficient for υ to satisfy the pivotal voting constraints and yield the same outcome as would occur under complete information *ex ante*. We therefore have to check the pivotal message constraints, given $\lambda \leq [2g-1]/2g$.

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 μ_{-i}. Suppose i has signal $s_i = 0$. Given (μ_{-i}, υ) 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

$$E[u(z, y, \mathbf{s})|0, 0, \boldsymbol{\mu}_{-i}, \upsilon, \mathbf{mpiv}(0, -1)] \geq E[u(z', y, \mathbf{s})|0, -1, \boldsymbol{\mu}_{-i}, \upsilon, \mathbf{mpiv}(0, -1)].$$

Given (μ, υ), i is message 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

$$\Pr(y^*|s_i, \mu_{-i}, \upsilon, \mathbf{mpiv}(m, m')) =$$

$$\frac{\Pr(\mathbf{mpiv}(m, m')|\boldsymbol{\mu}_{-i}, \upsilon, y^*) \Pr(y^*|s_i)}{\Pr(\mathbf{mpiv}(m, m')|\mu_{-i}, \upsilon, y^*) \Pr(y^*|s_i) + \Pr(\mathbf{mpiv}(m, m')|\mu_{-i}, \upsilon, x^*) \Pr(x^*|s_i)}$$

where

$$\begin{aligned}\Pr(\mathbf{mpiv}(0,-1)|\mu_{-i},\upsilon,y^*) &\equiv [\frac{1}{4}(1-f)^2+\frac{1}{2}f^2g(1-g)+\frac{1}{2}f(1-g)(1-f)],\\ \Pr(\mathbf{mpiv}(0,-1)|\mu_{-i},\upsilon,x^*) &\equiv [\frac{1}{4}(1-f)^2+\frac{1}{2}f^2(1-g)g+\frac{1}{2}fg(1-f)].\end{aligned}$$

Since $\Pr(y^*|s_i=0)=1/2$, i is willing to send $m_i=0$ rather than $m_i'=-1$ only if

$$\begin{aligned}\lambda &\leq \frac{1-2\Pr(y^*|0,\mu_{-i},\upsilon,\mathbf{mpiv}(0,-1))}{2[1-\Pr(y^*|0,\mu_{-i},\upsilon,\mathbf{mpiv}(0,-1))]}\\ &= \frac{f(1-f)(2g-1)}{[(1-f)^2+2f^2(1-g)g+2fg(1-f)]}\\ &< [2g-1]/2g.\end{aligned}$$

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 (μ,υ), i is message 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

$$\Pr(\mathbf{mpiv}(1,-1)|\mu_{-i},\upsilon,y^*)\equiv$$

$$[(1-f)^2+\frac{1}{2}f^2g(1-g)+\frac{1}{4}f(1-g)(1-f)+\frac{1}{4}fg(1-f)],$$

and

$$\Pr(\mathbf{mpiv}(1,-1)|\mu_{-i},\upsilon,x^*)\equiv$$

$$[(1-f)^2+\frac{1}{2}f^2(1-g)g+\frac{1}{4}fg(1-f)+\frac{1}{4}f(1-g)(1-f)].$$

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

$$\begin{aligned} \lambda &\leq \frac{1 - 2\Pr(y^*|1, \mu_{-i}, \upsilon, \mathbf{mpiv}(1, -1))}{2[1 - \Pr(y^*|1, \mu_{-i}, \upsilon, \mathbf{mpiv}(1, -1))]} \\ &= \frac{(2g-1)}{2g}. \end{aligned}$$

Therefore the binding message pivot constraint is that on the uninformed individual, in which case there exists a fully revealing debate equilibrium if and only if

$$\lambda \leq \frac{f(1-f)(2g-1)}{[(1-f)^2 + 2f^2(1-g)g + 2fg(1-f)]}.$$

Defining the left side of this inequality to be $\lambda(f, g)$ and observing $\lambda(f, g) \in (0, [2g-1]/2g)$ completes the proof. $\square$

6 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.

Austen-Smith, D. and T. Feddersen. 2004. Deliberation and voting rules. Forthcoming in D. Austen-Smith and J. Duggan (eds) *Social Choice and Strategic Decisions: Essays in Honor of Jeffrey S. Banks* Heidelberg: Springer.

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*.

Coughlan, P. 2001. In defense of unanimous jury verdicts: mistrials, communication and strategic voting. *American Political Science Review* 94:375-93.

Diermeier, D. and T. Feddersen. 2000. Information and congressional hearings. *American Journal of Political Science* 44:51-65.

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

Dryzek, J. and C. List. 2002. Social choice theory and deliberative democracy: a reconciliation. Forthcoming, *British Journal of Political Science.*

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

Eriksen, E. 2001. Democratic or technocratic governance? *NYU School of Law, Jean Monnet Center WP 6/01, Comment 6.*

Gerardi, D. and L. Yariv. 2004. Putting your ballot where your mouth is: an analysis of collective choice with communication. *Working Paper, Department of Economics, Yale University.*

Gilboa, I. and N. Vieille. 2004. Majority vote following a debate. *Social Choice and Welfare,* 23:115-126.

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

Hafer, C. and D. Landa. 2003. Deliberation as self-discovery. *Working Paper, Department of Politics, New York University.*

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

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

Shapiro, I. 2002. Optimal deliberation. *Journal of Political Philosophy*, 10:1-16.

Walker, J. and D. Lane 1994. The jury system. *Victorian Council for Civil Liberties, Criminal Justice* DP 25/02/94