A Note on Preference Uncertainty and Communication in Committees[1]

David Austen-Smith

MEDS, Kellogg Graduate School of Management

Northwestern University

Evanston, IL 60208

Tim Feddersen

MEDS, Kellogg Graduate School of Management

Northwestern University

Evanston, IL 60208

July 2008

Abstract

The recent formal literature on communication in committees indicates that truthtelling in committee deliberations is more likely under nonunanimous voting rules when individuals' private interests, or biases, are not common knowledge. The result, however, does not imply that such bias uncertainty is necessarily welfare improving in some appropriate sense. This Note suggests that bias uncertainty is in fact welfare improving and, further, argues that even when committee members prefer that biases be shared, there need not exist any cheap talk equilibria in which individuals can credibly reveal such information.

Introduction

It is well understood that when committee members vote under incomplete information, the resulting committee decision need not reflect the decision that would have been made under fully shared information (eg Austen-Smith and Banks 1996; Feddersen and Pesendorfer 1998). But people in committees often talk before voting and so have an opportunity to share decision-relevant information. Over the past few years there has been a growing strategic game-theoretic literature concerned to understand better what implications such communication might have for the character and quality of collective decisions under incomplete information.[1] One issue here concerns how different voting rules for reaching a final collective choice influences the information that might be shared in any prior debate.[2] Our focus in this paper is not on voting rules but rather on the normative consequences of preference uncertainty for the quality of collective choices under incomplete information. It is useful first, however, to review briefly the motivating positive results on the implications of such uncertainty for information sharing in debate.

To fix ideas, consider the canonical example of a jury that has to vote over whether to convict or acquit a defendant. The guilt or innocence of the defendant is subject to uncertainty. Each juror privately observes an informative but noisy signal from a given set of possible signals regarding the innocence or guilt of the defendant. If the defendant is truly guilty, then the probability of any juror receiving a signal suggesting as much is strictly greater than the probability the juror receives a signal suggesting innocence, although there is some likelihood of such an innocent signal is received; and similarly with respect to innocent signals conditional on the defendant being truly innocent. Jurors also have possibly different attitudes (biases) regarding just how much such information suffices to convince them of guilt. For example, if there are only two possible signals, some juror might require all committee members to have received the "guilty" signal before they are willing to convict, whereas another could be willing to convict on the basis of only one such signal among the jurors. As with their

signals, individuals' biases are private information. The committee uses some q-rule to make the final decision, where a q-rule is a voting rule under which conviction is chosen if and only if it receives at least q votes, where q is at least a majority of the jury. The jury may or may not deliberate prior to voting, where deliberation consists of jurors making costless (so-called cheap talk) speeches regarding any decision-relevant information they may have (in particular, their private signals).

In a very general setting of the sort described above, Coughlan (2000) demonstrates that full (collective decision-relevant) information sharing under complete information about committee members' biases is impossible under any q-rule unless all individuals have identical full information ordinal preferences over the alternatives. Austen-Smith and Feddersen (2006) go on to show that Coughlan's result is sensitive to the full information assumption on committee members' preferences: so long as the rule is not unanimity rule (i.e. q is less than the total number of committee members) and there is uncertainty about both individuals' preferences and signals, then there always exist some circumstances under which full information sharing of signals in debate is possible, thus inducing the same decision as would be made were all signals revealed simultaneously to all jurors. On the other hand, if the jury or committee uses a unanimity voting rule then preference uncertainty no longer provides any possibility for full information sharing in debate; there is always some type of individual with a strict incentive to dissemble in debate and, therefore, the veracity of any information conveyed in individuals' speeches is generally suspect.

The key intuition underlying the results is that bias uncertainty admits the possibility that any speaker belongs to the de facto winning coalition under complete information, a possibility that provides an incentive for truth-telling in debate. In contrast, when all individuals' biases are common knowledge prior to any debate, then those in any losing coalition under full information have no incentive to share information that improves the chances of the decision that would be most preferred under full information by a winning coalition (see also Meirowitz 2004, 2007).

By themselves, the preceding results do not imply that unanimity rule is in general inferior to any other q-rule. It is possible that unanimity rule performs better on average than other rules despite the fact that full information sharing in debate is (at least, on grounds of strategic rationality) impossible. More generally, an important problem is to identify the optimal rule for committee choice under incomplete information with debate.[3] This problem is particularly challenging if we insist that individuals' voting strategies are undominated conditional on the realized debate, a requirement that opens up the possibility, as illustrated by results sketched above, that voting rules affect the incentives for deliberation.[4]

In this paper we address a more limited welfare question that concerns the extent to which the de facto committee decision under a given voting rule with incomplete information and debate reflects the committee's decision conditional on all private information being common knowledge at the time of the vote (e.g. McLennan 1998; Feddersen and Pesendorfer 1997). Equivalently, given the rule, the goal is to maximize the aggregate expected payoff of those in the full information winning coalition. And the intuition suggested by the results above is that, at least for nonunanimous q-rules, bias uncertainty can improve the welfare of the full information winning coalition by facilitating more information sharing in debate: as already observed, when individuals are not sure ex ante whether they are members of the full information winning coalition, then they have an incentive to reveal private information in debate that is absent when they are confident they are not members of this coalition.

Unfortunately, even restricting attention to q-rules, addressing the latter welfare issue directly is complicated by having to describe the full equilibrium set to any more or less general incomplete information (debate and voting) game induced by the rule (see for example Austen-Smith and Feddersen (2005) or Dorazelski, Gerardi and Squintani (2003) on this issue). At this stage, no such characterization is available.[5] So rather than attempt a general result here, we instead explore the welfare implications of bias uncertainty in the context of much simplified committee structure predicated on the observation that when strategically rational individuals are choosing a best response strategy for information sharing in debate, they condition on

3

the event that their particular speech is *message pivotal*; that is, that if they say one thing then the resultant committee vote leads to one alternative whereas a different speech induces a committee decision in favour of the competing alternative. In effect, therefore, it is as if the speaker were addressing the (possibly hypothetical) pivotal member of the committee whose preferences (bias) and information on the relative value of the two alternatives is private information to that pivot. Thus the simplification is to reduce the problem to a game between two individuals, one of whom is a representative of a non-decisive coalition (the advisor) and the other a represenatative of a decisive coalition (the decision-maker).

The model

Consider a committee of two people in which one fixed individual has the right to make the committee decision between a fixed pair of alternatives $\{X, Y\}$. Each individual $i = 1, 2$ has private information $(b_i, s_i) \in \{x, y\} \times \{x, y\}$, where b_i is a preference parameter, or *bias*, and s_i is a noisy but informative *signal* regarding the alternatives. Let $(\mathbf{b}, \mathbf{s}) = ((b_1, b_2), (s_1, s_2))$ denote a profile of realized biases and signals. Assume signals are uncorrelated with biases. Let the prior probability on each signal $s \in \{x, y\}$ be $1/2$ and suppose that, all i, for distinct $i, j \in \{1, 2\}$

$$\Pr[s_i = s_j | s_i] = p \in (1/2, 1).$$

Assuming $p \in (1/2, 1)$ insures signals are both informative and noisy. For each $i = 1, 2$, suppose that while the prior probability that $b_i = x$ is $1/2$, realized bias types are correlated; that is, for distinct $i, j \in \{1, 2\}$

$$\Pr[b_i = b_j | b_i] = \rho \in [0, 1].$$

If $\rho = 0$ they have conflicting preferences with probability 1 whereas if $\rho = 1$ both have identical preferences. As suggested earlier, the idea here is that the decision-maker is a proxy for the pivotal voter defining the winning coalition under a q-rule in a larger committee, and the

other individual (the advisor) is unsure whether she too is a member of that winning coalition. Hereafter, let individual $i = 1$ be the advisor and individual $i = 2$ be the decision-maker.

Write $u(Z, b, \mathbf{s})$ for an individual's payoff from the collective choice $Z \in \{X, Y\}$, given that individual's bias is $b \in \{x, y\}$ and the profile of signals is $\mathbf{s} \in \{x, y\}^2$. Assume

$$
\begin{aligned}
u(X, x, \mathbf{s}) &= \begin{cases} 0 & if \quad \mathbf{s} = (y, y) \\ 1 & otherwise \end{cases} \\
u(Y, x, \mathbf{s}) &= 1 - u(X, x, \mathbf{s}) \\
u(X, y, \mathbf{s}) &= \begin{cases} 1 & if \quad \mathbf{s} = (x, x) \\ 0 & otherwise \end{cases} \\
u(Y, y, \mathbf{s}) &= 1 - u(X, y, \mathbf{s})
\end{aligned}
$$

Thus, a bias type $b = x$ (respectively, $b = y$) strictly prefers alternative X (respectively, Y) unless both signals constitute evidence for Y (respectively, X).

We consider two communication protocols. In the first protocol there is a single "debate" stage in which the advisor makes a cheap talk speech regarding her signal, following which the decision-maker chooses X or Y. Without any real loss of generality, we assume the available messages are simply the signals, $\{x, y\}$. For the second protocol, we add an earlier debate stage in which both players can declare their biases (again, through cheap talk speeches) following which the advisor sends a message regarding her signal as before and, finally, the decision-maker makes a decision. Note that, because an individual's payoffs, given her bias, depend exclusively on the final decision and the profile of realized signals $\mathbf{s}$, there can be no value in signaling both bias and signal simultaneously or of having a debate on signals prior to making any statements regarding biases. Thus the role of permitting bias revelation early in any committee discussion is to improve the possibility of coordination between committee members.

Begin with assuming only a single message stage in which the advisor makes a speech about her signal. Strategies are as follows: $\sigma : \{x, y\}^2 \to \Delta\{x, y\}$ is a debate strategy with $\sigma(b, s)$ being the probability that the advisor (individual 1) with bias b and signal s declares "$s = x$";

and $v : \{x, y\}^3 \rightarrow \Delta\{X, Y\}$ is the decision-maker's decision (vote) strategy with $v(b, s, n)$ being the probability that the decision-maker (individual 2) with bias b and signal s who hears a message $n \in \{x, y\}$ from the advisor chooses X. Let $\mu(s, \sigma) \subseteq \{x, y\}$ be the set of messages sent with strictly positive probability when signal s is observed.

Definition 1 *A message strategy σ is* fully revealing *if $\mu(x, \sigma) \cap \mu(y, \sigma) = \varnothing$.*

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

A *fully revealing debate equilibrium (FRDE)* for the one-stage debate protocol is a Perfect Bayesian Equilibrium (σ, v) such that σ is fully revealing. Hereon, we economize on notation by writing σ_s^b for the probability an advisor with bias type b and signal s sends message x; and writing $v_s^b(n)$ for the probability the decision-maker with bias b, signal s who hears message n chooses X.

Proposition 1 *There exists an FRDE iff $\rho \geq (1-p)$. If $\rho < (1-p)$ the advisor cannot reveal any information in equilibrium and the decision-maker chooses with his signal.*

Proof: Without loss of generality, we can suppose that in a fully revealing equilibrium we have, for all i, $\sigma_x^b = 1 - \sigma_y^b = 1$ for every $b \in \{x, y\}$. In this case the best response strategy for a decision-maker with bias b who observes signal s' and hears message s from the advisor is

$$v_{s'}^b(s) = \begin{cases} 1 & if & (s, s') = (x, x) \ \forall b \\ 1 & if & (s, s') \in \{\{x, y\}, \{y, x\}\} \text{ and } b = x \\ 0 & & otherwise \end{cases}$$

The incentive compatibility conditions for a FRDE are obviously satisfied for the decision-maker since he is correctly choosing as if he were fully informed. So, we only need to check the IC constraints for the advisor.

Without loss of generality, assume the advisor has bias $b_1 = x$. Then it is trivially the case (mod inversions of natural language) that she surely reveals a signal $s_1 = x$. So suppose she has observed $s_1 = y$. There are only two pivotal events in which the advisor's message affects the outcome here: either the decision-maker shares the advisor's bias (x) and has observed a signal $s_2 = y$; or the the decision-maker has the opposite bias (y) and has observed a signal $s_2 = x$. In the former event, the advisor's best response is to tell the truth and reveal $s_1 = y$; in the latter event, the advisor's best response is to lie and claim $s_1 = x$. The conditional probability of these two events is, respectively, ρp and $(1 - \rho)(1 - p)$. Therefore the sender is willing to tell the truth when she has signal y iff

$$\rho p \geq (1 - \rho)(1 - p);$$

that is, iff $\rho \geq (1 - p)$. If $\rho < (1 - p)$, then clearly the advisor has a strict best response to announce "$s = b$" irrespective of her signal and no information is credibly conveyed in debate (the decision-maker can always reveal his signal in debate, but this is of no consequence for the final outcome). Hence the decision-maker must choose on the basis of his own information. Suppose, without loss of generality, that $b_2 = x$. Then player 2 has a dominant strategy to choose X conditional on $s_2 = x$. If $s_2 = y$, however, his payoff from choosing X is $(1 - p)/2$ whereas that from choosing Y is $p/2$. Since $p > 1/2$, the best response is to choose Y. $\square$

It is worth emphasising that a FRDE can exist here for very small values of ρ, depending on just how informative is the signal, p: the more information that any given signal provides, the less concern the advisor exhibits about the probability of not being a member of the de facto 'winning coalition'.[6] This result is a direct analogue of the conditions defining existence of a FRDE for the case of q-rules in a committee of more than two individuals (Austen-Smith

and Feddersen, 2005, 2006), discussed informally in the Introduction, and shares the same fundamental intuition (see also the Appendix to this paper). In particular, as made explicit in the argument for the result, when deciding on her best message the sender conditions on being message pivotal and there are (in this model) two such events: in one the decision-maker shares the advisor's bias and signal; in the other the decision-maker has the opposite bias and signal. Although the distributions of bias and signal are quite independent, the realization of biases and signals are linked for the advisor through the event of being message pivotal. And although which event is relatively more likely depends symmetrically here on both the informativeness of the signal, p, and the likelihood of biases being the same, ρ, in general, if ρ is not too extreme the fact that a signal is informative shifts weight to the event in which the advisor has an incentive to tell the truth. The analytical advantage of the two-person model studied here, however, is that, at least up to the mixed equilibrium for the nongeneric event $\rho = (1 - p)$, the proposition completely describes the equilibrium set. Thus unlike committees with more than two persons and a q-rule, the welfare implications of bias uncertainty in this instance are easy to identify.

Proposition 2 *Assume the most informative equilibrium is played for every parameterization (p, ρ). (a) If $\rho > (1-p)$ then, ex ante, the decision-maker (respectively, advisor) strictly prefers biases to remain secret (respectively, revealed) before the game is played. (b) If $\rho < (1 - p)$ then, ex ante, both players strictly prefer biases to be revealed prior to the game being played.*

Proof: Suppose that both players' biases are made common knowledge prior to the game being played. Then either they share the same bias, in which case the most informative equilibrium is a FRDE, or they have opposing biases in which case (as is easy to confirm) no information is revealed by the advisor, player 1. The payoffs to the two players are therefore:

$$EU_1^{rev} = EU_2^{rev} = \rho + (1 - \rho) \left(\frac{1+p}{2} \right).$$

To see this, first note that if biases are the same, all information is revealed by player 1 and both players are assured a payoff of one. Next, suppose the biases are revealed to be distinct. Then no information is revealed and the decision-maker can do no better than choose according to his signal; that is, choose X iff $s_2 = x$. In this case the sender (player 1) receives one iff either $s_1 = s_2$, which occurs with probability p, or her bias is x (respectively, y) and $s_2 = x$ (respectively, y), which occurs with probability $(1-p)/2$. These facts yield the expression for EU_1^{rev}. Similarly, given the biases are distinct, the decision-maker obtains a payoff equal to one iff either $s_1 = s_2$ or her bias is x (respectively, y) and $s_2 = x$ (respectively, y). This justifies EU_2^{rev}.

Now suppose both players' biases remain secret. If $\rho > (1-p)$ then there exists a FRDE and the decision-maker obtains a payoff $EU_2^{pvt} = 1$ surely. The advisor, however, receives payoff one surely only if the biases are the same; if biases are different, then the advisor reveals her signal and receives payoff one iff $s_1 = s_2$. Thus $EU_1^{pvt} = \rho + (1-\rho)p$. On the other hand, if $\rho < (1-p)$ then the only equilibrium involves no information revelation and the decision-maker chooses according to his signal. Hence, following the reasoning for the case in which biases are revealed, we obtain

$$EU_1^{pvt} = EU_2^{pvt} = \frac{1+p}{2}.$$

Therefore, if $\rho > (1-p)$ the decision-maker gains $(1-\rho)(1+p)/2$ when biases are secret but the advisor loses $(1-\rho)(1-p)/2$. And when $\rho < (1-p)$, both players strictly prefer biases to be revealed. $\square$

If the likelihood that the advisor is in fact a member of the 'winning coalition', in that she shares a common bias with the decision-maker, is sufficiently large then the presence of bias uncertainty strictly improves the welfare of that coalition (i.e. the decision-maker) since all decision-relevant information can be shared in equilibrium. Because the two players agree about the desirability of bias-revelation when ρ is sufficiently low relative to p, however, it is interesting to ask whether providing an opportunity to coordinate directly by revealing biases

prior to debating signals can improve committee performance.

As remarked earlier, the definition of a fully revealing message strategy says nothing about whether or not the speaker reveals her bias b_i in debate; all that matters is that the speaker's collective decision-relevant datum, her signal s_i, is unequivocally revealed. It is apparent, however, that if biases are revealed they have no effect on any listener's decision given all signals are shared. The only reason for being concerned about a speaker's bias is that such knowledge, or lack thereof, influences a listener's beliefs about whether or not he or she is in the full information winning coalition: if a committee member knows that they are a member of the majority in a committee before any discussion of signals, then that individual's assessment of the likely message pivotal events is surely affected and, therefore, so is her willingness to share collective decision-relevant information prior to voting. Thus if revealing information regarding personal bias is relevant to the collective decision, then it must be that biases are revealed prior to the discussion of signals. So we suppose there are two cheap-talk debate stages prior to voting: in the first, individuals send a message regarding their biases, following which there the second debate stage involves individuals sending cheap-talk messages about their signals.

Strategies for the two-stage debate protocol are as follows. Let $\beta_i^{b,s} \in [0,1]$ denote the probability that individual i with bias b and signal s announces her bias is x; given bias-stage messages $(m, m') \in \{x, y\}^2$; let $\sigma_s^b(m, m') \in [0,1]$ is the probability individual 1, the advisor, with bias b and signal s who has announced her bias is m and heard a message that the other player's bias is m', announces that her signal is x; and let $v_s^b((m, m'), (n, n'))$ be the probability the decision-maker with bias b and signal s who announces bias m and signal n, hears the advisor's bias message m' and signal message n', chooses outcome X.

Proposition 3 *Assume each player has observed his or her particular bias and signal. If and only if $\rho \geq 1/2$, there exists an equilibrium in which each player truthfully reveals their bias in the first message round and (1) if the biases are the same, players truthfully reveal their*

10

signal in the second round and the receiver chooses on the basis of full information; (2) if the biases are different, players simply announce their true bias independently of their signal and the decision-maker chooses with his signal.

Proof: Consider the following strategies (specified only for an x-biased individual; those for a y-biased individual are symmetric):

$$\beta_i^{x,x} = \beta_i^{x,y} = 1, \ i = 1, 2$$

$$\sigma_x^x(m, m') = 1, \forall(m, m')$$
$$\sigma_y^x(m, x) = 1 - \sigma_y^x(m, y) = 0$$

$$v_x^x((m, m'), (n, n')) = 1, \forall((m, m'), (n, n'))$$
$$v_y^x((m, m'), (\cdot, x)) = v_y^x((x, y), (x, x)) = 1, \forall m$$
$$v_y^x((m, m'), (n, n')) = 0 \ otherwise$$

To confirm that this strategy profile constitutes an equilibrium if $\rho \geq 1/2$, we calculate

$$E[U_i^{x,x}|\beta_i^{x,x} = 1, \cdot] = E[U_i^{x,x}|\beta_i^{x,x} = 0, \cdot] = \rho + (1 - \rho)\frac{(1+p)}{2}$$
$$E[U_i^{x,y}|\beta_i^{x,y} = 1, \cdot] = \rho + (1 - \rho)\frac{(1+p)}{2}$$
$$E[U_i^{x,y}|\beta_i^{x,y} = 0, \cdot] = \rho\frac{(1+p)}{2} + (1 - \rho)$$

where $E[U_i^{b,s}|\beta_i^{b,s}, \cdot]$ denotes the equilibrium expected payoff for individual i with bias b and signal s from adopting bias-debate strategy $\beta_i^{b,s}$. Hence telling the truth at the bias revelation stage is incentive compatible iff $E[U_i^{x,y}|\beta_i^{x,y} = 1, \cdot] \geq E[U_i^{x,y}|\beta_i^{x,y} = 0, \cdot]$, which obtains if $\rho \geq 1/2$. And given biases are revealed truthfully, the subsequent debate and decision strategies are easily checked to be best responses. $\square$

Thus allowing an opportunity to coordinate through sharing bias information before revealing anything about decision-relevant signals, cannot improve the welfare properties for the decision-maker when $\rho < 1 - p$.[7] One might also consider the possibility of an alternative communication protocol for the case $\rho < 1 - p$ allowing, for example, mediated communication. A natural extension of this model might take a mechanism design approach. While a full blown mechanism design analysis is beyond the scope of this paper it is easy to see that no mechanism can yield outcomes equivalent to a FRDE. As in Proposition 1, the sender who has observed a signal different from his bias has an incentive to misrepresent his signal.

Recall that the stylized committee model assumes that the advisor is a representative agent for a minority coalition while the decision-maker stands in for a majority coalition. In the Appendix we consider an extension of the model in which agents are unsure whether they are part of a majority or a minority. In such a setting there is no comparable equilibrium in an extension of the model to a three or more person committee with majority rule. That is, as demonstrated in the Appendix, the possibility of bias revelation seems peculiar to the advisor (known minority)/decision-maker (known majority) committee structure. Equilibrium bias revelation followed by (not necessarily fully) informative signal sharing is generally impossible when a committee uses a two-stage debate protocol followed by majority rule voting to determine the committee choice.

Note that, so long as $\rho \geq 1/2$ in the identified equilibrium (say, the *bias revelation* equilibrium), both individuals reveal their biases in the first round of talk; if their biases are the same then all information is revealed in the second round whereas, if biases are different, no further decision-relevant information sharing takes place. It is immediate, therefore, that the expected payoff to any (b, s)-type individual from playing the bias revelation equilibrium identified in Proposition 3 equals the expected payoff they achieve if biases could be revealed ex ante by some external agent. Specifically, for all i, b, s,

$$E[U_i^{b,s} | \text{ biases revealed}] = \rho + (1 - \rho)\frac{(1 + p)}{2}.$$

Furthermore, because $p > 1/2$, Propositions 1 and 3 together imply there is also a FRDE with a bias debate stage when $\rho \geq 1 - p$: both individuals babble during the first round of talk and, subsequently, all decision-relevant information is revealed during the second round of communication exactly as described in Proposition 1; that is, for all i, b, s, $\beta_i^{b,s} = 1/2$ and $\sigma_x^b = 1 - \sigma_y^b = 1$. Hence we have the same welfare comparisons as described in Proposition 2, with the further observation that, when $\rho \geq 1/2 > 1 - p$, the advisor strictly prefers to play the bias revelation equilibrium to playing the FRDE with no bias revelation (of course, the decision-maker has the opposite preferences).

In sum, bias uncertainty can improve the welfare of the de facto full information 'winning coalition'. And even in those circumstances in which bias revelation is available in equilibrium, there exists a superior (from the decision-maker's perspective) equilibrium in which biases remain private information.

Conclusion

There is clearly a great deal left to be learned regarding communication in committees. In particular, identifying the optimal voting rule in the presence of debate, and comparing this rule with the optimal rule without communication, are important and open questions. However, they have as yet proved largely intractable if we insist that voting be undominated conditional on the realized messages in debate. On the other hand, the discussion here provides further support for the intuition that full knowledge of committee members' underlying preferences, or biases, coming into a decision-making process is a deterrent to credible and useful sharing of information germane to the collective decision. Moreover, the expected quality of any committee choice, at least as reflected by the likelihood that any such decision coincides with the decision that would be made under fully shared information, is greater with bias uncertainty than otherwise.

Appendix

In this Appendix we present a variation of the two-person committee model of the text, involving a three-person committee, $i = 1, 2, 3$, using majority rule, $q = 2$, to make final decisions.[8] We first provide conditions for a FRDE and then argue that, unlike the two-person case, there exists no equilibrium in which all individuals can share their biases credibly in debate to facilitate coordination and subsequently discuss individuals's signals regarding the state, $\mathbf{s} = (s_1, s_2, s_3)$.

The committee consists of three people, $i = 1, 2, 3$, that has to choose between a fixed pair of alternatives $\{X, Y\}$. Exactly as for the two-person case, each individual has private information regarding their bias and a signal about the true state, $(b_i, s_i) \in \{x, y\} \times \{x, y\}$; let $(\mathbf{b}, \mathbf{s}) = ((b_1, b_2, b_3), (s_1, s_2, s_3))$ denote a profile of realized biases and signals. As before, assume signals are uncorrelated with biases and suppose $\{A_1, A_2\}$ are states such that, for all i,

$$\Pr[s_i = x | A_1] = \Pr[s_i = y | A_2] = p \in (1/2, 1)$$

with the common prior probability on A_1 being $1/2$; and, for all i, assume $\Pr[b_i = x] = r \in (0, 1)$.

Naturally extending the two-person model, an individual's payoff from the collective choice $Z \in \{X, Y\}$, given that her bias is $b \in \{x, y\}$ and the profile of signals is $\mathbf{s} \in \{x, y\}^3$, is given by:

$$u(X, x, \mathbf{s}) = \begin{cases} 0 & if \quad \mathbf{s} = (y, y, y) \\ 1 & otherwise \end{cases}$$
$$u(Y, x, \mathbf{s}) = 1 - u(X, x, \mathbf{s})$$
$$u(X, y, \mathbf{s}) = \begin{cases} 1 & if \quad \mathbf{s} = (x, x, x)\} \\ 0 & otherwise \end{cases}$$
$$u(Y, y, \mathbf{s}) = 1 - u(X, y, \mathbf{s})$$

The committee's choice is determined by a majority vote. As observed in the text, there is

no equilibrium in which all individuals vote informatively without any pre-vote communication. To see there can be a FRDE for some environments, assume all individuals truthfully reveal their signals; that is, using the notation of the text extended in the obvious manner to the three-person setting, $\sigma_i(b, x) = 1 - \sigma_i(b, y) = 1$ for all i and biases b. Then there is fully shared information at the voting stage and the unique undominated voting equilibrium conditional on these messages is for each individual i to vote 'sincerely' for their most preferred alternative as defined by $(b_i, (s_1, s_2, s_3))$. Clearly, any x-biased individual with an x signal, or y-biased individual with y signal, has a dominant strategy to reveal their signal truthfully. So, without loss of generality, consider an x-biased individual i with a y signal: $(b_i, s_i) = (x, y)$. Then there are essentially two events in which i is message pivotal:

(1) both of the other committee members are x-biased and both have observed a y signal, which occurs with (conditional) probability

$$\frac{1}{2} r^2 \left(p^3 + (1-p)^3 \right);$$

(2) both of the other committee members are y-biased and both have observed an x signal, which occurs with (conditional) probability

$$\frac{1}{2}(1-r)^2 \left((1-p)^2 p + (1-p)p^2 \right) = \frac{1}{2}(1-r)^2 p(1-p).$$

In event (1), individual i strictly prefers to tell the truth while, in event (2), i strictly prefers to dissemble. Making the requisite calculations, telling the truth in debate conditional on all others fully revealing their signals is incentive compatible for an x-biased individual with a y signal if and only if

$$\left(\frac{r}{1-r} \right)^2 \geq \frac{p(1-p)}{(1-p)^3 + p^3}.$$

Similarly, incentive compatibility of truthtelling for a y-biased individual with an x signal is insured if

$$\left(\frac{1-r}{r} \right)^2 \geq \frac{p(1-p)}{(1-p)^3 + p^3}.$$

15

For every $r \in (0, 1)$, therefore, there exists some $p < 1$ for which there exists an equilibrium in which all individuals reveal their signals truthfully in debate and subsequently vote under full information to yield the ex post first best de facto majority preferred outcome. Although certainly more complicated than the formula of Proposition 1 in the text, the intuition and comparative statics of this result are identical.

Proposition 3 in the text showed that sharing information about biases prior to an influential debate regarding signals is possible in equilibrium if the two individuals of the committee are sufficiently likely to share the same bias; i.e. the advisor is sufficiently likely to be a member of the winning coalition, defined by the decision-maker (pivot). This turns out not to be a general result with respect to committee size and decision rule.

As before, suppose there are two cheap-talk debate stages prior to voting: in the first, individuals send a message regarding their biases, following which there the second debate stage involves individuals sending cheap-talk messages about their signals. Although the notation is perhaps unnecessary to establish the result of interest here, to avoid ambiguity in discussion, it seems sensible to specify formal strategies for the protocol. Specifically: $\beta_i^{b,s} \in [0, 1]$ is the probability that individual i with bias b and signal s announces her bias is x; given bias-stage messages $\mathbf{m} = (m_1, m_2, m_s) \in \{x, y\}^3$; let $\sigma_i^{b,s}(m, \mathbf{m}_{-i}) \in [0, 1]$ is the probability individual i with bias b and signal s who has announced her bias is m and heard messages $\mathbf{m}_{-i}$ regarding the other players' biases, announces that her signal is x; and similarly, let $v_i^{b,s}((m, \mathbf{m}_{-i}), (n, \mathbf{n}_{-i}))$ be the probability this individual i chooses outcome X given the history of speeches in debate.

Proposition 4 *Assume each player has observed his or her particular bias and signal. There exists no equilibrium in which each player truthfully reveals their bias in the first message round and all members of the revealed majority subsequently reveal their signals prior to voting.*

Proof: Suppose, to the contrary, that all individuals reveal their biases during the first debate stage and, in the second debate stage, those committee members sharing the majority bias

16

reveal their signals while those in the minority offer no further credible information; all individuals then vote for their most preferred alternative conditional on their information sets at the time. In particular, members of the revealed majority vote identically but, assuming some heterogeneity in biases across the committee, not under complete information; the minority individual on the other hand, should she exist, votes under complete information. (It is easy to check that, conditional on full bias revelation at the first stage, these subsequent stage message and vote strategies constitute equilibrium play.) Formally, the strategies (specified only for an x-biased individual; those for a y-biased individual are symmetric) are described by:

$$\beta_i^{x,x} = \beta_i^{x,y} = 1, \ i = 1, 2, 3$$

$$\sigma_i^{x,x}(\mathbf{m}) \ = \ 1, \forall \mathbf{m}$$
$$\sigma_i^{x,y}(m_i, x, x) \ = \ 1 - \sigma_i^{x,y}(m_i, y, y) = 0$$

$$v_i^{x,x}(\mathbf{m}, \mathbf{n}) \ = \ 1, \forall (\mathbf{m}, \mathbf{n})$$
$$[\exists j \neq i : m_i = m_j \ \& \ n_j = x] \ \implies \ v_i^{x,y}(\mathbf{m}, (n_i, \mathbf{n}_{-i})) = 1$$
$$v_i^{x,y}((x, (y, y)), (x, x, x)) \ = \ 1$$
$$v_i^{x,y}(\mathbf{m}, \mathbf{n}) \ = \ 0 \ \textit{otherwise.}$$

We argue that this profile cannot be an equilibrium. To this end, suppose the specified strategy profile constitutes an equilibrium and consider an x-biased individual, say $i = 1$, with a y signal; now suppose that i deviates from the prescribed strategy and lies about her bias. There are three cases, depending on the revealed (truthfully, by supposition) biases of the other individuals, $2, 3$.

Case 1: both of the others credibly reveal they are y-biased.

Given the declared bias profile and the specified continuation strategies, $i = 1$ is message-pivotal at the second debate stage when both have x signals: $s_2 = s_3 = x$. In this case,

individual 1 has a best response to lie regarding her signal, else Y is chosen surely at the voting stage. Given i's revealed bias is y, her second stage debate message is credible to the (true) majority of $\{2, 3\}$ who therefore have a best voting decision in favour of X. However, since both $2, 3$ declare their x-signals in the second stage and signals are informative, this coalition would have unanimously chosen X had individual 1 truthfully declared her bias to be x in the first stage, revealed herself as the minority and subsequently played the specified strategy in the second debate stage. Thus, in this case, individual 1 gains no advantage by lying about her bias in the first debate stage.

Case 2: both of the others credibly reveal they are x-biased.

In this case, i is message-pivotal when both 2 and 3 have y-signals. Hence, $i = 1$'s undominated best response is to tell the truth about her signal, $s_1 = y$, since the only circumstance in which she strictly prefers the Y decision is when all three signals are y. But having lied about her bias in the first stage and declared a y bias, any message in favour of Y at the second stage is ignored by the majority as not credible. However, even so, the majority votes together for Y because, given the two credible messages that at least two (informative) y signals have been observed, voting for Y is a best response. As for Case 1, therefore, $i = 1$ cannot profit from deviating in the first debate stage and lying about her bias.

Case 3: the others credibly reveal they have opposing biases.

Given that $i = 1$ has lied about her bias, the apparent majority in the committee is y-biased: without loss of generality, assume $m_2 = b_2 = x$ and $m_3 = b_3 = y$. Consequently, at the second debate stage, the (apparent) minority individual 2 speaks in favour of X irrespective of his signal, and the (apparent) majority coalition member 3 both reveals her signal credibly. Moreover, all three individuals believe the second stage debate messages from individuals 1 and 3 and subsequently vote using this information. Therefore, 1's best response message at the second stage is to reveal her signal $n_1 = s_1 = y$ truthfully and subsequently vote for X irrespective of the y-biased individual 3's message. Following such a second stage debate,

18

individual 3 has (given $b_3 = y$ and 1's revealed y signal) a best response decision to vote for Y. In turn, this makes individual 2, the apparent minority member of the committee, vote pivotal in which case his best response is to vote for Y if and only if $s_2 = s_3 = y$. Hence, given the others follow the prescribed strategies, the deviation from full truthtelling regarding her bias is strictly profitable individual $i = 1$ in this case.

In sum, because Cases 1, 2 and 3 are exhaustive, an x-biased individual with a y signal can never be made worse off by deviating to reveal a y bias in the first debate stage, and can be make strictly better off in one case. Thus the supposition that the specified strategy profile constitutes and an equilibrium in undominated strategies is false. $\square$

Notes

[1] For example, Austen-Smith (1990); Calvert and Johnson (1998); Coughlan (2000); Doraszelski et al (2003); Gerardi and Yariv (2007); Meirowitz (2006, 2007); Hafer and Landa (2007); Austen-Smith and Feddersen (2005, 2006); Caillaud and Tirole (2006).

[2] See Austen-Smith and Feddersen (2005), Gerardi and Yariv (2007) and Meirowitz (2007) for a discussion of voting rules and information aggregation.

[3] Chwe (2006) derives the optimal voting rule when there is no debate. Interestingly, this rule turns out to be nonmonotonic in votes and is thus not a q-rule.

[4] In a very general framework, Gerardi and Yariv (2007) show that all non-unanimous q rules are equivalent in that the sets of sequential equilibrium outcomes induced by any such rules are identical once voting is preceded by deliberation. However, the assumptions supporting this result do not insist that voting strategies are undominated conditional on the realized debate.

[5] Likewise, the optimal mechanism here is as yet unavailable in general. Chwe's (2006) result for majority voting with incomplete information suggests such a mechanism will be complex.

[6] We note too that even though there is no reason for the decision-maker to reveal his signal in debate, there is also no reason here for him not to do so credibly, a fact that motivates our equilibrium terminology.

[7] Of course, full bias revelation in debate is possible if all subsequent messages regarding individuals' signals are ignored. But this case is clearly equivalent to the situation with no debate at all.

[8] The basic example to follow was introduced in Austen-Smith and Feddersen (2006) where the conditions for a FRDE were identified. For completeness, we summarize this finding here.

References

Austen-Smith, David. 1990a. "Information transmission in debate." *American Journal of Political Science* 34 (February): 124-52.

Austen-Smith, David and Jeffrey S. Banks. 1996. "Information aggregation, rationality, and the Condorcet Jury Theorem." *American Political Science Review* 90 (March): 34-45.

Austen-Smith, David and Timothy J. Feddersen. 2005. "Deliberation and voting rules." In *Social Choice and Strategic Decisions: Essays in Honor of Jeffrey S. Banks,* eds. David Austen-Smith and John Duggan. Heidelberg:Springer.

Austen-Smith, David and Timothy J. Feddersen. 2006. "Deliberation, preference uncertainty and voting rules." *American Political Science Review* 100 (May): 209-17.

Caillaud, Bernard and Jean Tirole. 2006. "Consensus building: how to persuade a group" CNRS Working Paper.

Calvert, Randall L. and James Johnson. 1998. "Rational argument, political argument and democratic deliberation." Presented at the Annual Meeting of the American Political Science Association.

Chwe, Michael. 2006. "A robust and optimal anonymous procedure for Condorcet's model." UCLA, Working Paper.

Coughlan, Peter J. 2000. "In defense of unanimous jury verdicts: mistrials, communication and strategic voting." *American Political Science Review* 94 (June): 375-93.

Doraszelski, Ulrich, Dino Gerardi and Francesco Squintani. 2003. "Communication and voting with double-sided information." *Contributions to Theoretical Economics* 3(1), Article 6.

Feddersen, Timothy J. and Wolfgang Pesendorfer. 1997. "Voting behavior and information aggregation in large elections with private information." *Econometrica* 65(5): 1029-58.

Feddersen, Timothy J. and Wolfgang Pesendorfer. 1998. "Convicting the innoncent: the inferiority of unanimous jury verdicts." *American Political Science Review* 92 (March): 23-35.

Gerardi, Dino and Leeat Yariv. 2007. "Deliberative Committees." *Journal of Economic Theory* 134(1): 317-338.

Glazer, Jacob and Ariel Rubinstein. 2001. "Debates and decisions: on a rationale of argumentation rules." *Games and Economic Behavior* 36(2):158-73.

Hafer, Catherine and Dimitri Landa. 2007. "Deliberation as self-discovery and the institutions for political speech." forthcoming in *Journal of Theoretical Politics.*

McLennan, Andrew. 1998. "Consequences of the Condorcet Jury Theorem for beneficial information aggregation by rational players." *American Political Science Review.* 92(2):413-18.

Meirowitz, Adam. 2004. "In defense of exclusionary deliberation: communication and voting with private beliefs and values" forthcoming in *Journal of Theoretical Politics.*

Meirowitz, Adam. 2007. "Deliberative democracy or market democracy: designing institutions to aggregate preferences and information." *Quarterly Journal of Political Science* 1(4): 373-92.

Ottaviani, Marco and Peter Sorensen. 2001. "Information aggregation in debate: who should speak first?" *Journal of Public Economics* 81(3):393-422.