Deliberation and epistemic democracy

Huihui Ding, Marcus Pivato

Deliberation and epistemic democracy*

Huihui Ding and Marcus Pivato†

February 23, 2021

Abstract

We study the effects of deliberation on epistemic social choice, in two settings. In the first setting, the group faces a binary epistemic decision analogous to the Condorcet Jury Theorem. In the second setting, group members have probabilistic beliefs arising from their private information, and the group wants to aggregate these beliefs in a way that makes optimal use of this information. During deliberation, each agent discloses private information to persuade the other agents of her current views. But her views may also evolve over time, as she learns from other agents. This process will improve the performance of the group, but only under certain conditions; these involve the nature of the social decision rule, the group size, and also the presence of "neutral agents" whom the other agents try to persuade.

JEL classification: D71, D83.
Keywords: Deliberation; Epistemic social choice; Condorcet Jury Theorem; Probabilistic belief aggregation; Multiplicative pooling

Just as a banquet to which many contribute dishes is finer than a single plain dinner, for this reason in many cases a crowd judges better than any single person.

—Aristotle, *Politics* III.11 1286a

1 Introduction

In many collective decisions, there is an objectively correct answer, and the group wants to find it. For example, a criminal trial jury must decide whether a defendant is innocent or guilty. The Supreme Court must determine the constitutional validity of laws or lower-court decisions. The directorate of the Central Bank must evaluate the risks of inflation and recession over the coming year. The senior management of a firm must determine which business strategy will maximize long-term profits. Finally, blue ribbon commissions and

*Supported by CHOp (ANR17-CE26-0003) and Labex MME-DII (ANR11-LBX-0023-01). Ding gratefully acknowledges the support of the CY Inititive (CODEX). We thank Pierpaolo Battigalli, Richard Bradley, David Budescu, Alain Chateauneuf, Julio Davila, Franz Dietrich, Simona Fabrizi, Maria Gallego, Mario Roberto Gilli, Ariane Lambert-Mogiliansky, Antoine Mandel, John McCoy, Klaus Nehring, Drazen Prelec, Peter Sørensen, Emily Tanimura, and Bauke Visser for helpful comments. We also thank two referees for their constructive suggestions.

†THEMA, CY Cergy Paris Université, huihui.ding@cyu.fr and marcuspivato@gmail.com

scientific committees must advise policy-makers on questions of scientific fact. *Epistemic social choice theory* studies the conditions under which voting rules or opinion aggregation methods can deliver accurate answers to such questions (see Pivato, 2013, 2017 or Dietrich and Spiekermann, 2020a,b for summaries). This has inspired a parallel literature in political philosophy on *epistemic democracy* (Cohen, 1986; Landemore and Elster, 2012; Schwartzberg, 2015; Goodin and Spiekermann, 2018). Starting with the Condorcet Jury Theorem, most models in epistemic social choice theory have assumed that the opinions of different voters are stochastically independent, conditional on the true state of nature.[1] But in reality, the opinions of voters are correlated, because they deliberate with one another. Indeed, there is now an extensive literature on *deliberative democracy* which argues that deliberation should *improve* the epistemic competency of groups (Fishkin and Laslett, 2008; Landemore, 2013; Landemore and Page, 2015; Estlund and Landemore, 2018). But it is not clear that deliberation is always beneficial in this regard. Deliberation enables agents to pool information, but they might double-count this information (Berg, 1997). Deliberation can also lead to "groupthink" (Janis, 1972; Solomon, 2006; Visser and Swank, 2007; Mayo-Wilson et al., 2013), informational cascades (Banerjee, 1992; Bikhchandani et al., 1992, 1998), and other pathologies. On the balance, does deliberation really lead to more epistemically reliable group decisions?

Suppose that any initial heterogeneity of beliefs amongst the agents arises from heterogeneity of private information. And suppose that during deliberation, every agent truthfully reveals *all* of her private information to the group, and every other agent correctly understands this information and updates her beliefs accordingly. Then deliberation will generally improve the reliability of the group, because the collective decision will fully incorporate the pooled information of all group members. So if deliberation were this simple, then the answer to the question in the previous paragraph would be trivially affirmative. But in reality, things are not so simple, for two reasons. First, agents may either lie or withhold information to manipulate the group decision to their own advantage. Second, even amongst honest agents, communication can be costly: it may take time and effort for an agent to clearly and credibly convey her private information to other group members so that they fully understand it and update their beliefs accordingly. Indeed, if "private information" is interpreted in the broadest sense, to encompass all of an agent's prior education and professional experience, then it is simply infeasible for her to communicate *all* her information to the group; she must select some small subset to disclose. Over the past two decades, a considerable literature on *strategic deliberation* has developed in response to the first problem (see Section 5 for a review). But there has been very little examination of the second problem.[2] This paper aims to fill this gap.

We will construct a stylized model of deliberation with costly communication amongst Bayesian agents, and investigate whether deliberation improves the reliability of the group's

[1] An early exception is Ladha (1992, 1995), who was the first to extend the Condorcet Jury Theorem to correlated voters. For more recent related work, see Pivato (2017) and the references cited there.

[2] Exceptions are Fishman and Hagerty (1990) and Glazer and Rubinstein (2001, 2004), who consider models of persuasion and argumentation in which agents can only disclose a subset of their information, due to time constraints or other communication costs. Landa (2019) also discusses the importance of information overload in deliberation.

decision. In our model, the true state of the world is unknown; the group wants to determine this state. The agents have each received a set of private signals ("evidence") that are informative about the true state. In each time period, each piece of evidence is either *private* or *public*. *Private* evidence is known by only one (or a few) agents. *Public* evidence is known by everyone. Agents are Bayesian; at any moment in time, each agent's beliefs are obtained by combining her prior beliefs with her currently available evidence —that is, her own private evidence and all the publicly available evidence. Deliberation is a way for agents to "disclose" some of their private evidence, turning it into *public* evidence, so as to modify the beliefs of the other agents.[3] But evidence disclosure takes time. An agent might have many pieces of private evidence, but she can only disclose one piece at a time. Thus, unlike most deliberation models in the literature, our model is *diachronic*; we track the evolution of each agent's beliefs (and evidence base) over time.

Furthermore, evidence disclosure is costly: communication involves an expense of time and effort, both for the sender and the receivers. So an agent will not disclose her private evidence without an incentive. Her incentive is to convince the other agents of the correctness of her current views. Unlike the literature on strategic deliberation, we assume that the agents deliberate in *good faith* —that is, each agent simply wants the group decision to converge to the correct answer (or at least, what she currently *believes* to be the correct answer). Strategic deliberation is typically driven by heterogeneity of preferences. In our model, there is no heterogeneity of preferences —only heterogeneity of *beliefs*. Also, unlike the literature on strategic deliberation, agents in our model do not harbour second-order beliefs about other agents' undisclosed private information (and third-order beliefs about other agents' second-order beliefs, and so on). They only have first-order beliefs about the state of the world, based upon the evidence currently available to them.

Agents take turns disclosing single pieces of evidence. Each agent only discloses evidence that will move the beliefs of other group members closer to her own current beliefs. But at the same time, her own beliefs also evolve, as a consequence of the evidence revealed by other agents. This process continues until the group reaches a *deliberative equilibrium* in which every agent is either unwilling or unable to disclose further evidence. This raises two questions. First, do such equilibria exist? Second, how accurate are the resulting group decisions? To answer these questions, we will consider three stylized scenarios.

Case 1. Suppose that the group must make a simple binary decision by majority vote (as in the Condorcet Jury Theorem). Thus, at any stage in the deliberation, the goal of each agent is that a *majority* of the other agents agree with her. If she is already part of the majority, then this is automatically true. Otherwise, if her current opinion is in the minority, then she will present new evidence to the group (if she has any) to try to persuade voters in the opposing majority to agree with her.

Case 2. Now suppose the informed agents cannot vote, but instead act as "advisors" to an external decision-maker ("the President"). Each advisor wants the President

[3] This is similar to the *hidden-profiles paradigm* of Stasser and Titus (1985), which has played a prominent role in the social psychology literature on deliberation; see Lu et al. (2012) and Maciejovsky and Budescu (2019) for recent reviews.

to adopt her opinion. The President is initially uninformed but Bayesian, so her decision will be determined by the balance of *publicly available* evidence. So each advisor wants the publicly available evidence to support her current opinion.

Case 3. Finally, suppose that the group is no longer faced with a binary decision; instead, it must select a *probability distribution* over two or more possibilities. Each agent's probabilistic beliefs are determined by her prior probability and her current evidence, via Bayes rule. The goal of each agent is thus to present evidence so that the probabilistic beliefs of other group members converge to her own current beliefs.

In *Case 1*, suppose that at least one member of the group begins with a *neutral opinion*, either because she has no private information, or because her private information is equally balanced between positive and negative evidence. Then deliberation often leads to the same outcome as would be achieved if (counterfactually) all the agents revealed *all* their private evidence (Theorems 1 and 2). Likewise, in *Case 2*, under certain conditions, deliberation leads the President to make an optimally informed decision (Theorems 3 and 4). Finally, in *Case 3*, under certain conditions, deliberation yields the decision that would occur under full information pooling (Theorems 5 and 6).

The rest of the paper is organized as follows. Section 2 presents the model. Section 3 considers deliberation in a binary decision as in Cases 1 and 2 above, and contains Theorems 1 to 4. Section 4 considers deliberation in the context of probability aggregation, as in Case 3 above, and contains Theorems 5 and 6. Section 5 reviews prior literature and Section 6 discusses the hypotheses of the model. All proofs are in the Appendices.

2 Framework

Let $\mathcal{X}$ be a finite set of possible states of the world. Let $\mathcal{Y}$ be a space of possible signal values. Let $\Delta^*(\mathcal{X})$ be the set of all probability measures on $\mathcal{X}$ with *full support* —i.e. such that every element of $\mathcal{X}$ receives nonzero probability. Define $\Delta^*(\mathcal{Y})$ in the same way. Let $\rho : \mathcal{X} \longrightarrow \Delta^*(\mathcal{Y})$ be a function; if the true state of the world is $x \in \mathcal{X}$, then the signals will be conditionally independent, identically distributed (i.i.d) random variables drawn from the probability distribution $\rho(x)$, which we will write as ρ_x.

Let $\widetilde{x}$ be the (unknown) true state of the world. Suppose a Bayesian agent has prior beliefs about $\widetilde{x}$ given by $\pi \in \Delta^*(\mathcal{X})$, and she receives a sequence $\mathbf{y} = (y_1, y_2, \ldots, y_M)$ of i.i.d. random signals drawn from $\rho_{\widetilde{x}}$. Let $B_{\mathbf{y}}$ be her posterior beliefs about $\widetilde{x}$, given $\mathbf{y}$. Then $B_{\mathbf{y}}$ is the following probability distribution over $\mathcal{X}$:

$$B_{\mathbf{y}}(x) \quad = \quad \frac{1}{\text{(SNC)}}\, \pi(x) \cdot \prod_{m=1}^{M} \rho_x(y_m), \qquad \text{for all } x \in \mathcal{X}. \tag{2A}$$

Here and throughout the paper, "(SNC)" refers to "Some Normalization Constant", needed to ensure that the expression in question defines a probability distribution. These normal-

ization constants are not important to the analysis, so we will not specify them explicitly.[4]

Now let $\mathcal{I}$ be a set of agents. For all $i \in \mathcal{I}$, let $\pi_i \in \Delta^*(\mathcal{X})$ be a probability distribution describing i's prior beliefs, before acquiring any information. Let $\mathcal{M}$ be a finite indexing set, and let $\{y_m\}_{m\in\mathcal{M}}$ be a set of i.i.d. random signals drawn from ρ_x, where x is the (unknown) true state of the world. We refer to these signals as *evidence*. For all $i \in \mathcal{I}$, let $\mathcal{M}_i \subset \mathcal{M}$ be the set of evidence received by agent i —this is i's initial private information. (We do not necessarily assume that these sets are disjoint.) Meanwhile, let $\mathcal{C}^0 \subset \mathcal{M}$ be the set of evidence that is common information at time 0. (We assume it is disjoint from the sets $\mathcal{M}_i$.) Thus, before deliberation, formula (2A) says that i's initial beliefs are given by

$$B_i^0(x) \quad = \quad \frac{\pi_i(x)}{(\text{SNC})} \prod_{c\in\mathcal{C}^0} \rho_x(y_c) \cdot \prod_{m\in\mathcal{M}_i} \rho_x(y_m), \qquad \text{for all } x \in \mathcal{X}. \tag{2B}$$

We can assume without loss of generality that

$$\mathcal{M} \quad = \quad \mathcal{C}^0 \sqcup \bigcup_{i\in\mathcal{I}} \mathcal{M}_i, \tag{2C}$$

because any evidence which is not in this union will never be learned by the group after any amount of deliberation, and is therefore irrelevant to our analysis.

Deliberation takes place in a series of "rounds", indexed by the set of natural numbers $\mathbb{N} = \{0, 1, 2, \ldots\}$. For any $t \in \mathbb{N}$, let $\mathcal{C}_i^t \subseteq \mathcal{M}_i$ be the part of agent i's evidence which she has disclosed (i.e. made into common information) at time t. Thus, $\mathcal{P}_i^t := \mathcal{M}_i \setminus \mathcal{C}_i^t$ is the part of agent i's evidence which remains "private" at time t. Let

$$\mathcal{C}^t \quad := \quad \mathcal{C}^0 \sqcup \bigcup_{i\in\mathcal{I}} \mathcal{C}_i^t. \tag{2D}$$

This is the set of evidence that is common information at time t. Let B_i^t be the probabilistic beliefs of agent i at time t. This is a combination of her prior, the publicly available evidence, and her own (undisclosed) private evidence. By applying (2A), we obtain:

$$B_i^t(x) \quad = \quad \frac{\pi_i(x)}{(\text{SNC})} \cdot \prod_{p\in\mathcal{P}_i^t} \rho_x(y_p) \cdot \prod_{c\in\mathcal{C}^t} \rho_x(y_c), \qquad \text{for all } x \in \mathcal{X}. \tag{2E}$$

In other words, agent i's beliefs about the world at time t are entirely determined by the evidence which is available to her at time t. In particular, she does not speculate about the possibility of further, undisclosed evidence still held by other agents.[5] From formula (2E) it is clear that, when agent i discloses evidence (i.e. transfers it from $\mathcal{P}_i^t$ to $\mathcal{C}^t$), her own beliefs will not change. However, this disclosure may affect the beliefs of other people. Conversely, agent i's own beliefs can evolve as $\mathcal{C}^t$ accumulates more and more evidence from

[4] To be precise, in equation (2A), $(\text{SNC}) = \sum_{x\in\mathcal{X}} \pi(x) \left(\prod_{m=1}^{M} \rho_x(y_m) \right)$.

[5] We discuss this assumption further after Example (a) in Section 3, and also in Section 6.

other people. At any moment in time, her deliberation behaviour (i.e. the sort of evidence she discloses) depends on her current beliefs. So as her beliefs evolve, her behaviour may change. She may stop advocating one position (i.e. stop disclosing evidence supporting this position), and even start advocating a position she previously opposed. In Sections 3 and 4, we will explore two special cases of this model.

3 Deliberating binary decisions

In this section, we will focus on the simplest nontrivial example of the model, where $\mathcal{X} = \mathcal{Y} = \{\pm 1\}$. Fix $p \in (\frac{1}{2}, 1)$, and suppose that $\rho(x|x) = p$ and $\rho(-x|x) = 1-p$, for both $x \in \mathcal{X}$. This is essentially the set-up of the Condorcet Jury Theorem. We will describe $+1$-valued signals as "positive" and -1-valued signals as "negative". We assume all agents have a common prior π, which assigns probability $\frac{1}{2}$ to each state.

For all $i \in \mathcal{I}$ and $t \in \mathbb{N}$, the *opinion* of agent i at time t is the variable $s_i^t \in \{-1, 0, 1\}$ describing whether i believes the positive state or the negative state to be more probable, according to the probabilistic beliefs B_i^t defined by equation (2E). Formally,

$$s_i^t \quad := \quad \begin{cases} 1 & \text{if} \quad B_i^t(1) > B_i^t(-1); \\ 0 & \text{if} \quad B_i^t(1) = B_i^t(-1); \\ -1 & \text{if} \quad B_i^t(1) < B_i^t(-1). \end{cases} \tag{3A}$$

Using equation (2E) and the common prior π, it is easily verified that s_i^t is entirely determined by the amount of positive and negative evidence available to i at time t. Formally,

$$s_i^t \quad = \quad \text{sign}\left(\sum_{p \in \mathcal{P}_i^t} y_p + \sum_{c \in \mathcal{C}^t} y_c \right). \tag{3B}$$

Dissent and disclosure. We assume that it is easy for agents to learn their peers' opinions during each round of deliberation, simply by asking them yes/no questions or holding a straw vote. But in this section, we assume that it is *not* possible for agents to learn the underlying beliefs of their peers. (See Section 4 for a model with belief disclosure.) Furthermore, it is difficult to learn what evidence —or even *how much* evidence —their peers have to justify these opinions. We have abstractly represented each "piece of evidence" as a single binary signal. But in reality, it may by a complex corpus of facts, analysis, interpretation and arguments that may take considerable time and effort for an agent to clearly explain to her peers. Thus, an agent will not "disclose" this evidence (i.e. explain these facts and arguments) unless she has an incentive to do so. We assume that each agent wants the collective decision to be correct. At any time during deliberation, she believes that her *current* opinion is correct; thus, she seeks to persuade other group members to agree with her current opinion.[6] She will disclose evidence only if it advances

[6]See Section 6 for further discussion of this assumption. Also, note that $s_i^t = 0$ does not mean that i has "no opinion" —it means that i thinks ± 1 are *equally probable*, and hence the correct answer for the group is to remain ambivalent.

this goal. To be precise, she will only disclose evidence if her current opinion *disagrees* with the current collective decision —in this case, we say that she ***dissents*** from the group. Agents who already agree with the current collective decision will not disclose evidence, because such disclosure is costly, and they have no reason to incur this cost.

Although it will not play any formal role in the model that follows, it might be helpful to rationalize these behavioural assumptions with a stylized utility function. Let $0 < \epsilon < 1$. Suppose agent i starts with M_i pieces of private evidence, and during the course of deliberation, she discloses C_i of them. Suppose that the group decision is $x \in \{-1, 0, 1\}$, but agent i's opinion is y. Then i's final utility will be $-|x - y| - \epsilon\, C_i/M_i$. So disclosing each piece of evidence is costly, but this cost is outweighed by i's desire for the group decision to agree with her opinion. Thus, at any time t, she is always willing to disclose more evidence, *if* she thinks it will increase by more than ϵ the probability of the group agreeing with her. If ϵ is very small, then even a tiny gain in the probability of the group agreeing with her opinion is sufficient incentive for i to disclose more evidence. But if the group already agrees with her, then she will not incur the cost of disclosing further evidence.

Deliberative equilibrium. For most of this section, we will assume that the collective decision is made by *majority vote* (except in subsection 3.3). Formally, for any $t \in \mathbb{N}$, the ***majority opinion*** at time t is defined

$$\text{Maj}^t \quad := \quad \text{sign}\left(\sum_{i \in \mathcal{I}} s_i^t\right) \ \in \ \{-1, 0, 1\}, \tag{3C}$$

where $\{s_i^t\}_{i \in \mathcal{I}}$ are the individuals' opinions from formula (3B). Let $i \in \mathcal{I}$. If $s_i^t = \text{Maj}^t$, then we assume that i remains silent during round t. If $s_i^t > \text{Maj}^t$,[7] then we assume that i may disclose one piece of positive evidence from $\mathcal{P}_i^t$, so that it becomes part of $\mathcal{C}^{t+1}$. On the other hand, if $s_i^t < \text{Maj}^t$, then we assume that i may disclose one piece of negative evidence from $\mathcal{P}_i^t$, so it becomes part of $\mathcal{C}^{t+1}$.

Note that we only assume that voter i *may* disclose evidence in round t —not that she *will* disclose evidence. Whether or not i *actually* discloses evidence in round t depends on two things: first, whether she has any pertinent evidence left to disclose (see Example (a) below), and second, the nature of the deliberation protocol. In the *parallel* protocol discussed in Section 3.2 below, every dissenting agent can disclose evidence during round t (if they have any). But in the *serial* protocol of Section 3.1, only *one* dissenting agent can disclose evidence during each round. This distinction has implications for the outcome of deliberation, as we shall see later.

After each disclosure, everyone in the group will update their beliefs based on the newly revealed evidence. This process of deliberation must end in finite time (because $\mathcal{C}^1 \subset \mathcal{C}^2 \subset \mathcal{C}^3 \subset \cdots \subset \mathcal{C}^T \subseteq \mathcal{M}$ and $\mathcal{M}$ is finite). When it ends, all the agents are silent, either because they have no evidence left to disclose, or because they have no incentive to disclose their remaining evidence. We then say that the group is in a ***deliberative equilibrium***.

[7] This occurs if $s_i^t = 1$ and $\text{Maj}^t \in \{-1, 0\}$, or if $s_i^t \in \{0, 1\}$ and $\text{Maj}^t = -1$.

It might seem that in any such equilibrium, there will be unanimous consensus. The next example shows that this is not the case.

Example (a) Suppose $\mathcal{I} = \{h, i, j, k, \ell\}$, with $\mathcal{M}_h = \{\oplus_1\}$, $\mathcal{M}_i = \{\oplus_2, \oplus_3, \ominus_1\}$, $\mathcal{M}_j = \{\ominus_2, \ominus_3, \ominus_4, \ominus_5\}$, $\mathcal{M}_k = \{\ominus_6, \ominus_7, \ominus_8, \ominus_9\}$ and $\mathcal{M}_\ell = \{\ominus_{10}, \ominus_{11}, \ominus_{12}, \ominus_{13}\}$, where the signals $\oplus_1, \oplus_2, \oplus_3$ are positive while $\ominus_1, \ldots, \ominus_{13}$ are negative. Also, suppose $\mathcal{C}^0 = \emptyset$. Thus,

$$\begin{aligned}\sum_{m\in\mathcal{M}_h} y_m &= \sum_{m\in\mathcal{M}_i} y_m = 1, \quad \text{so that } s_h^0 = s_i^0 = 1, \text{ while}\\ \sum_{m\in\mathcal{M}_j} y_m &= \sum_{m\in\mathcal{M}_k} y_m = \sum_{m\in\mathcal{M}_\ell} y_m = -4, \quad \text{so that } s_j^0 = s_k^0 = s_\ell^0 = -1.\end{aligned}$$

The collective decision is made by majority vote. The initial majority decision is -1. Agents h and i dissent from this majority, so they disclose their positive evidence to try to convince j, k and ℓ. Suppose that during the first three rounds of deliberation, h and i disclose signals $\oplus_1$, $\oplus_2$, and $\oplus_3$. Then at time $t = 3$, we have $\mathcal{C}^3 = \{\oplus_1, \oplus_2, \oplus_3\}$, $\mathcal{P}_h^3 = \emptyset$, $\mathcal{P}_i^3 = \{\ominus_1\}$, $\mathcal{P}_j^3 = \{\ominus_2, \ldots, \ominus_5\}$, $\mathcal{P}_k^3 = \{\ominus_6, \ldots, \ominus_9\}$ and $\mathcal{P}_\ell^3 = \{\ominus_{10}, \ldots, \ominus_{13}\}$. Thus,

$$\begin{aligned}\sum_{c\in\mathcal{C}^3} y_c + \sum_{p\in\mathcal{P}_h^3} y_p &= 3, \quad \text{and} \quad \sum_{c\in\mathcal{C}^3} y_c + \sum_{p\in\mathcal{P}_i^3} y_p = 3-1 = 2, \quad \text{while}\\ \sum_{c\in\mathcal{C}^3} y_c + \sum_{p\in\mathcal{P}_j^3} y_p &= \sum_{c\in\mathcal{C}^3} y_c + \sum_{p\in\mathcal{P}_k^3} y_p = \sum_{c\in\mathcal{C}^3} y_c + \sum_{p\in\mathcal{P}_\ell^3} y_p = 3-4 = -1,\end{aligned}$$

so that $s_h^3 = s_i^3 = 1$ while $s_j^3 = s_k^3 = s_\ell^3 = -1$. In other words, h and i have failed to change the minds of j, k and ℓ, and they have now run out of positive evidence. Four agents still have undisclosed negative evidence, but none of them have any reason to disclose it; i will not disclose her negative evidence because it undermines her own (positive) opinion, while j, k and ℓ will not disclose their negative evidence because they already agree with the majority decision. Thus, the group is in deliberative equilibrium, but h and i still disagree with j, k and ℓ. ♢

A natural question arises in Example (a): Why do h and i not deduce from the negative opinions of j, k and ℓ that there is negative evidence that h and i currently do not know? The answer is that while h and i can deduce that j, k and ℓ know *some* negative information, they cannot deduce *how much* negative evidence they know. (Nor can h deduce how much *positive* information is known by i, or vice versa —at least, until deliberative equilibrium is reached.) From the perspective of h and i, it is possible that $\mathcal{M}_j = \{\ominus_2,\ominus_3,\ominus_4,\ominus_5,\ominus_6,\ominus_7,\oplus_4,\oplus_5\}$, $\mathcal{M}_k = \{\ominus_2,\ominus_3,\ominus_4,\ominus_5,\ominus_6,\ominus_7,\oplus_6,\oplus_7\}$, and $\mathcal{M}_\ell = \{\ominus_2,\ominus_3,\ominus_4,\ominus_5,\ominus_6,\ominus_7,\oplus_8,\oplus_9\}$, where $\oplus_4$, $\oplus_5$, $\oplus_6$, $\oplus_7$, $\oplus_8$ and $\oplus_9$ are six hypothetical positive signals (two for each of j, k and ℓ), while $\ominus_2$, $\ominus_3$, $\ominus_4$, $\ominus_5$, $\ominus_6$ and $\ominus_7$, are six hypothetical negative signals, which are unwittingly shared by all of j, k and ℓ. This hypothesis is consistent with the pattern of behaviour that h and i observe from j, k and ℓ during the entire deliberation process, up to and including the equilibrium. But if h and i believed this, then they would believe that j, k and ℓ hold six pieces of positive

evidence in total, and also six pieces of negative evidence, so it would be rational for h and i to maintain their own positive opinions throughout the deliberation (including the equilibrium), given their private information and what has been publicly disclosed.

Furthermore, once we enrich the model to allow agents to hold second-order beliefs about one another's unrevealed evidence, we could enrich it further to allow the agents to hold *third*-order beliefs about one another's second-order beliefs. For example, h and i might think that in fact $\mathcal{M}_j = \mathcal{M}_k = \mathcal{M}_\ell = \{\oslash_1, \oslash_2\}$, but also think that j, k and ℓ maintain their negative opinions because each of them believes (incorrectly) that the other two hold several other pieces of undisclosed negative evidence; this would be consistent with their behaviour during and after deliberation.

Indeed, in a model where agents can have second-order and higher-order beliefs, *any* pattern of deliberative behaviour would be consistent with some set of reasonable hypotheses on the part of the agents. One solution to this underdetermination problem would be a more detailed model, deploying stronger assumptions about each agents' knowledge of the other agents' information sources, so as to constrain their higher-order beliefs.[8] But such a model would be complex and not necessarily realistic. Instead, we opt for a simpler approach: our agents take their own evidence at face value, and do not speculate about the hidden evidence that other agents might possess. We do not suppose that agents are *unaware* that other agents may have undisclosed evidence, or that they lack a "theory of mind". But they do not form probabilistic second-order beliefs about other agents' possible undisclosed evidence (and third-order beliefs about other agents' second-order beliefs, etc.), because they lack the background information and cognitive resources needed to construct and consistently update such beliefs; see Section 6 for further discussion.[9]

For similar reasons, our model of deliberation is not a game; hence, a deliberative equilibrium is *not* a Nash equilibrium. A deliberative equilibrium is a stationary state in a diachronic process of nonstrategic sequential information disclosure. A diachronic model of *strategic* deliberation would be an extensive-form game; a subgame-perfect Nash equilibrium of such a game would require each agent at each time to anticipate the possible disclosures that other agents might make in the future, and to preemptively disclose her own evidence accordingly. In our model, agents do not hold beliefs about what private information other agents have, or even *how much* information other agents have, so it is not possible for them to engage in this sort of forward-looking strategic reasoning. (But see Appendix C for a simple game-theoretic interpretation of our results.)

Reliability and full disclosure. If the agents hold a majority vote at time t, then the decision is given by formula (3C). The *reliability* of this decision is the probability that it

[8] For example, in Aumann's (1976) model, each agent knows the information *partitions* of the other agents; see Sections 4 and 6 for further discussion on this point.

[9] A referee has suggested that, even if the agents cannot form precise probabilistic second-order beliefs about each other, they might cope with this uncertainty by maximizing the minimum expected utility over a *set* of probabilistic second-order beliefs, or using one of the other non-expected utility models investigated in decision theory. This is an interesting avenue for future investigation.

correctly identifies the true state.[10] We say there is *full disclosure* if all agents have complete information —that is $\mathcal{C}^t = \mathcal{M}$. In this case, all agents update to identical posterior beliefs via formula (3B); thus, the group is unanimous.

Proposition. *The (unanimous) majority decision under full disclosure achieves the maximum reliability possible given the information in* $\mathcal{M}$.

Since communication is costly, full disclosure may be unachievable. Fortunately, it may also be unnecessary. We will say that a deliberative equilibrium is *full-disclosure equivalent* if the collective decision in the equilibrium is the same as the consensus that *would be* reached under full disclosure. So it is sufficient to seek full-disclosure equivalence. For example, the equilibrium in Example (a) is full-disclosure equivalent, because the majority decision is negative, and likewise

$$\sum_{c\in\mathcal{C}^0} y_c + \sum_{m\in\mathcal{M}_h} y_m + \sum_{m\in\mathcal{M}_i} y_m + \sum_{m\in\mathcal{M}_j} y_m + \sum_{m\in\mathcal{M}_k} y_m + \sum_{m\in\mathcal{M}_\ell} y_m = 0+1+1-4-4-4 = -10 < 0.$$

Not all deliberative equilibria are full-disclosure equivalent. Also, the agents themselves might not *know* that their equilibrium is full-disclosure equivalent. (If h and i understood this in Example (a), then presumably they would switch to negative opinions.) But it is beyond the scope of this paper to model the agents' beliefs after deliberation ends.

Deliberation protocols. We have not yet described the timing of evidence disclosure —what we call the *protocol*. We will consider two possible protocols. In the *serial* protocol, only *one* dissenting agent can speak during each round of deliberation. In the *parallel* protocol, *all* dissenting agents can speak during each round of deliberation. The distinction between the two protocols is best understood as a question of the relative speed at which agents can transmit information, versus the speed at which they can incorporate new information into their beliefs. In the serial protocol, agents can incorporate new information very quickly: every time an agent reveals information, the other agents immediately update their beliefs. In the parallel protocol, agents update their beliefs more slowly. Thus, *all* dissenting agents have a chance to disclose new information, and then (perhaps during a brief interlude), all agents update their beliefs before the next round of deliberation begins.

This small difference in timing causes a big difference in the outcome. In the *serial* protocol, deliberation leads to full-disclosure equivalence under broad conditions (Theorem 1). But in the *parallel* protocol, deliberation is guaranteed to yield full-disclosure equivalence only under very special conditions (Theorem 2). These results depend on two assumptions.

(A) After some point during the deliberation, the agents have disjoint private information sets. That is, there is some $t \in \mathbb{N}$ such that $\mathcal{P}_i^t \cap \mathcal{P}_j^t = \emptyset$ for all distinct $i, j \in \mathcal{I}$.

[10] A collective decision of "0" is assigned reliability 0.5. Also, we neglect the possibility of strategic voting raised by Austen-Smith and Banks (1996) and Feddersen and Pesendorfer (1996, 1997, 1998, 1999).

(B) There is some agent $i \in \mathcal{I}$ who is *internally neutral*, by which we mean $\sum_{m \in \mathcal{M}_i} y_m = 0$.

To satisfy Assumption (B), it is sufficient (but not necessary) that $\mathcal{M}_i = \emptyset$. Meanwhile, Assumption (A) is satisfied if $\mathcal{M}_i \cap \mathcal{M}_j = \emptyset$ for all distinct $i, j \in \mathcal{I}$. But this is not necessary; it is sufficient that all private information shared by two or more agents be disclosed at *some* point during the deliberation process.

One reason why this is plausible is that evidence held by two or more agents is more likely to be disclosed early in deliberation, so that the group converges to a situation where Assumption (A) is satisfied. There are two informal arguments for this claim. First, if agents disclose their evidence at random, then a piece of evidence which is known by three people is three times as likely to be disclosed as a piece of evidence known only by one person. So shared evidence tends to get disclosed earlier. Second, recall that disclosure is costly: it takes time and effort to clearly and credibly communicate evidence to other agents. If some other agents already share this evidence, then this could reduce these costs —hence, this evidence is likely to be disclosed first.

It is beyond the scope of this paper to provide a formal justification for the informal arguments in the previous paragraph. But it is also not necessary. We do not claim that Assumption (A) and (B) will *always* be satisfied. They are not standing assumptions in our model, and they are not always satisfied in reality. They could easily be violated, as shown in Examples (b) and (c) below. Even if they *are* satisfied, the agents themselves might not know this, as in the discussion following Example (a). Our main results simply say that *if* these assumptions are satisfied, then deliberation will lead to a good outcome.

3.1 The serial protocol

For any $t \in \mathbb{N}$, let $\{s_i^t\}_{i \in \mathcal{I}}$ be the individuals' opinions from formula (3B). Let $\mathcal{I}_+^t$ be the set of agents with *positive* opinions at time t. Let $\mathcal{I}_-^t$ be the set of agents with *negative* opinions at time t. Let $\mathcal{I}_0^t$ be the set of agents with *neutral* opinions at time t. Let $E^0 := \sum_{c \in \mathcal{C}^0} y_c$ be the balance of public evidence at time zero. Suppose $E^0 \neq 0$; we will say there is *initial disagreement* if at least one agent has an opinion different from $\text{sign}(E^0)$ at time 0. In other words, if $E^0 > 0$, then $\mathcal{I}_-^0 \sqcup \mathcal{I}_0^0 \neq \emptyset$, whereas if $E^0 < 0$, then $\mathcal{I}_+^0 \sqcup \mathcal{I}_0^0 \neq \emptyset$. During each round of deliberation, exactly one dissenting agent will reveal exactly one piece of evidence. To be precise, during round t, if $\text{Maj}^t = 1$, then exactly *one* (randomly chosen) agent in $\mathcal{I}_-^t \cup \mathcal{I}_0^t$ reveals one piece of negative evidence. Likewise, if $\text{Maj}^t = -1$, then exactly one (randomly chosen) agent in $\mathcal{I}_+^t \cup \mathcal{I}_0^t$ reveals one piece of positive evidence. Finally, if $\text{Maj}^t = 0$, then exactly one (randomly chosen) agent in $\mathcal{I}_+^t \cup \mathcal{I}_-^t$ reveals one piece of evidence (either positive or negative). We have not specified the probability distributions by which these dissenting agents are "randomly chosen". If we specified these distributions, then we could describe serial protocol deliberation as a stochastic process. But it turns out that the precise probability distribution doesn't matter.

Theorem 1 *Assume (A) and (B), and suppose that either $E^0 = 0$, or there is initial disagreement. Then the serial protocol always reaches a deliberative equilibrium that is full-disclosure equivalent.*

The next three examples show why the hypotheses of Theorem 1 are needed.

Examples. (b) To see why Assumption (A) is required, suppose that $\mathcal{I} = \{0, 1, 2, 3, 4\}$ and $\mathcal{C}^0 = \emptyset$ (so that $E^0 = 0$). Suppose that $\mathcal{M}_0 = \emptyset$ (so agent 0 is internally neutral, in accord with Assumption (B)), while $\mathcal{M}_1 = \{\oplus_1, \oplus_1', \oslash_{12}, \oslash_{13}, \oslash_{14}\}$, $\mathcal{M}_2 = \{\oplus_2, \oplus_2', \oslash_{12}, \oslash_{23}, \oslash_{24}\}$, $\mathcal{M}_3 = \{\oplus_3, \oplus_3', \oslash_{23}, \oslash_{13}, \oslash_{34}\}$, and $\mathcal{M}_4 = \{\oplus_4, \oplus_4', \oslash_{14}, \oslash_{24}, \oslash_{34}\}$, where the signals $\oplus_1$, $\oplus_1'$, $\oplus_2$, $\oplus_2'$, $\oplus_3$, $\oplus_3'$, $\oplus_4$ and $\oplus_4'$ are positive, while $\oslash_{12}$, $\oslash_{13}$, $\oslash_{14}$, $\oslash_{23}$, $\oslash_{24}$ and $\oslash_{34}$ are negative. Thus, each of the four agents has three pieces of negative evidence and two pieces of positive evidence, so each agent has a negative opinion overall. So there is a unanimous negative consensus, and the group is already in deliberative equilibrium at time 0. However, there are eight pieces of positive evidence in total, and only six pieces of negative evidence. So the consensus after full disclosure would have been positive.

(c) To see why Assumption (B) is required, suppose $\mathcal{I} = \{i, j, k\}$, and $\mathcal{M}_i = \{\oplus_i, \oplus_i'\}$, $\mathcal{M}_j = \{\oplus_j\}$, $\mathcal{M}_k = \{\oplus_k, \oslash_k, \oslash_k'\}$, where $\oplus_i$, $\oplus_i'$, $\oplus_j$, and $\oplus_k$ are distinct positive signals, while $\oslash_k$ and $\oslash_k'$ are negative. Thus, Assumption (A) is satisfied, but Assumption (B) is violated. Also suppose $\mathcal{C}^0 = \emptyset$, so $E^0 = 0$. Thus, $s_i^0 = s_j^0 = 1$ while $s_k^0 = -1$, so that $\text{Maj}^0 = 1$. In Round 0, agent k dissents from this positive majority opinion, and discloses one piece of negative evidence. We now have $\mathcal{P}_i^1 = \{\oplus_i, \oplus_i'\}$, $\mathcal{P}_j^1 = \{\oplus_j\}$, $\mathcal{P}_k^1 = \{\oplus_k, \oslash_k\}$, and $\mathcal{C}^1 = \{\oslash_k'\}$; thus, $s_i^1 = 1$, $s_j^1 = 0$, and $s_k^1 = -1$, so $\text{Maj}^1 = 0$. Suppose that in Round 1, agent i dissents from the now-neutral majority opinion, and discloses one piece of positive evidence. We now have $\mathcal{P}_i^2 = \{\oplus_i\}$, $\mathcal{P}_j^2 = \{\oplus_j\}$, $\mathcal{P}_k^2 = \{\oplus_k, \oslash_k\}$, and $\mathcal{C}^2 = \{\oslash_k', \oplus_i'\}$; thus, $s_i^2 = 1$, $s_j^2 = 1$, and $s_k^2 = 0$, so $\text{Maj}^2 = 1$. In Round 2, agent k dissents from the new positive majority opinion, and discloses one piece of negative evidence. We now have $\mathcal{P}_i^3 = \{\oplus_i\}$, $\mathcal{P}_j^3 = \{\oplus_j\}$, $\mathcal{P}_k^3 = \{\oplus_k\}$, and $\mathcal{C}^3 = \{\oslash_k, \oslash_k', \oplus_i'\}$; thus, $s_i^3 = s_j^3 = s_k^3 = 0$, so that $\text{Maj}^2 = 0$. In other words, the group is in deliberative equilibrium in round 3. However, there are four pieces of positive evidence and only two pieces of negative evidence, so the full-disclosure decision would have been positive. Thus, this equilibrium is *not* full-disclosure equivalent.[11]

To see how Assumption (B) solves this problem, suppose we add an internally neutral agent n. In Round 0, agent k dissents and discloses negative evidence as before. Thus, $s_i^1 = 1$, $s_j^1 = 0$, and $s_k^1 = -1$, but now $s_n^1 = -1$ also, so that $\text{Maj}^1 = -1$. In Round 1, agents i and j dissent from this majority, so one of them will disclose positive evidence. This leads to two cases.

Case 1. If i discloses evidence in Round 1, then the majority opinion in Round 2 is neutral, as before. Thus, in Round 2, agent k dissents again. Thus, as before we have $\mathcal{P}_i^3 = \{\oplus_i\}$, $\mathcal{P}_j^3 = \{\oplus_j\}$, $\mathcal{P}_k^3 = \{\oplus_k\}$, and $\mathcal{C}^3 = \{\oslash_k, \oslash_k', \oplus_i'\}$; so that $s_i^3 = s_j^3 = s_k^3 = 0$. However, $s_n^3 = -1$, because n's opinion is determined by $\mathcal{C}^3$; thus, now $\text{Maj}^3 = -1$. Since i, j and k have neutral opinions, they dissent from this negative majority, and one of them will disclose her remaining positive evidence. Suppose it is i who discloses. (The argument

[11] In Round 1, another possibility is that agent k dissents again, and discloses another piece of negative evidence, leading to a *negative* majority in Round 2. In response, agent i dissents and discloses positive evidence in Round 2, so that in Round 3 the situation is the same as what was described above.

in the other two cases is very similar.) In this case, we have $\mathcal{C}^4 = \{\oslash_k, \oslash'_k, \oplus_i, \oplus'_i\}$ and $\mathcal{P}^4_i = \mathcal{P}^4_n = \emptyset$, while $\mathcal{P}^4_j = \{\oplus_j\}$ and $\mathcal{P}^4_k = \{\oplus_k\}$. Thus, $s^4_i = s^4_n = 0$ while $s^4_j = s^4_k = 1$. Thus, $\text{Maj}^4 = 1$. Agents i and n dissent from this majority, but they have no evidence left to disclose. Thus, the group reaches a deliberative equilibrium in Round 4.

Case 2. If j discloses evidence in Round 1, then $s^2_i = 1$, while $s^2_j = s^2_k = s^2_n = 0$, so that $\text{Maj}^2 = 1$. Thus, in Round 2, agent k once again dissents. We now have $\mathcal{C}^3 = \{\oslash_k, \oslash'_k, \oplus_j\}$ and $\mathcal{P}^3_j = \mathcal{P}^3_n = \emptyset$, while $\mathcal{P}^3_i = \{\oplus_i, \oplus'_i\}$ and $\mathcal{P}^3_k = \{\oplus_k\}$. Thus, $s^3_j = s^3_n = -1$, while $s^3_i = 1$ and $s^3_k = 0$, so that again, $\text{Maj}^3 = -1$. Agents i and k dissent from this negative majority, so one of them will disclose positive evidence. Suppose it is i (the argument for k is similar). Then $\mathcal{C}^4 = \{\oslash_k, \oslash'_k, \oplus_j, \oplus_i\}$ and $\mathcal{P}^4_j = \mathcal{P}^4_n = \emptyset$, while $\mathcal{P}^4_i = \{\oplus'_i\}$ and $\mathcal{P}^4_k = \{\oplus_k\}$. Thus, $s^4_j = s^4_n = 0$ while $s^4_i = s^4_k = 1$. Thus, $\text{Maj}^4 = 1$. Agents j and n dissent from this majority, but they have no evidence left to disclose. Once again, the group reaches a deliberative equilibrium in Round 4.

In either case, the group reaches a deliberative equilibrium that is full-disclosure equivalent. (Like Example (a), this shows that even a full-disclosure equivalent deliberative equilibrium does not require unanimous consensus amongst the agents.)

(d) Suppose $E^0 \neq 0$ and there is no initial disagreement. Then deliberative equilibrium is reached immediately, since there are no dissenting agents. But the resulting equilibrium is not necessarily full-disclosure equivalent. For example, suppose that $\mathcal{C}^0$ consists of exactly two pieces of positive evidence. Suppose $\mathcal{I} = \{n\} \sqcup \mathcal{I}^-$, where n is internally neutral (so that Assumption (B) is satisfied), while for all $i \in \mathcal{I}^-$, the set $\mathcal{M}_i$ consists of a single piece of negative evidence (and these sets are disjoint, so Assumption (A) is satisfied.) Then $\sum_{m\in\mathcal{M}_n} y_m + \sum_{c\in\mathcal{C}^0} y_c = 2 > 0$ so that $s^0_n = 1$, while for all $i \in \mathcal{I}^-$, we have $\sum_{m\in\mathcal{M}_i} y_m + \sum_{c\in\mathcal{C}^0} y_c = -1 + 2 > 0$ so that $s^0_i = 1$. But if $|\mathcal{I}^-| \geq 3$, then $\sum_{c\in\mathcal{C}^0} y_c + \sum_{i\in\mathcal{I}} \sum_{m\in\mathcal{M}_i} y_m \leq 2 - 3 < 0$, so full disclosure would yield a *negative* consensus opinion. So the equilibrium is not full-disclosure equivalent in this example. ♢

Theorem 1 assumes that all pieces of evidence are equally informative —i.e. they are all drawn from the same conditional probability distribution ρ, as explained at the start of Section 3. If some pieces of evidence are more informative than others, then deliberative equilibria might no longer be full-disclosure equivalent. Consider the following scenario. For all $m \in \mathcal{M}$, let $p_m \in (\frac{1}{2}, 1)$, and suppose that the signal y_m is drawn from the conditional distribution ρ_m, where $\rho_m(x|x) = p_m$ and $\rho_m(-x|x) = 1 - p_m$, for both $x \in \{\pm 1\}$. Thus, more informative signals correspond to larger values of p_m. Let $w_m = \log\left(\frac{p_m}{1-p_m}\right)$; then using equations (2E) and (3A) and the common prior $\pi = (\frac{1}{2}, \frac{1}{2})$, we obtain the following generalization of equation (3B):

$$s^t_i \quad = \quad \text{sign}\left(\sum_{p\in\mathcal{P}^t_i} w_p\, y_p + \sum_{c\in\mathcal{C}^t} w_c\, y_c\right).$$

Now suppose $\sum_{c\in\mathcal{C}^t} w_c\, y_c = 0.5$ for some $t \geq 1$, and there are ten opinionated agents, each with no remaining positive evidence, but each holding a piece of private negative evidence

m_i with $w_{m_i} = 0.2$. (Also suppose that Assumption (B) is satisfied.) The total weight of all this private negative evidence is -2. If all the private negative evidence was disclosed, then we would have $E = 0.5 - 2.0 = -1.5$, and the group consensus would be negative. *However*, individually each agent performs the mental computation $0.5 - 0.2 = 0.3$, and ends up with a *positive* personal opinion, so no one reveals any of their evidence, and the deliberative equilibrium yields a positive consensus; hence it is not full-disclosure equivalent.

3.2 The parallel protocol

In the serial protocol of Section 3.1, only one agent could disclose evidence during each round of deliberation. But in the parallel protocol, *every* agent can disclose evidence in each round. However, an agent will disclose evidence only if she dissents from the majority. For any $t \in \mathbb{N}$, let $\mathcal{N}^t := \{i \in \mathcal{I};\ \sum_{p \in \mathcal{P}_i^t} y_p = 0\}$. (Thus, $\mathcal{N}^0$ is the set of internally neutral agents who appear in Assumption (B).) We will require a third assumption.

(C) For all $t \in \mathbb{N}$, no agent in $\mathcal{N}^t$ discloses information in round t.

If $\mathcal{P}_n^t = \emptyset$ for all $n \in \mathcal{N}^t$, or if $\text{Maj}^t = \text{sign}\left(\sum_{c \in \mathcal{C}^t} y_c\right)$, then Assumption (C) is trivially satisfied at time t. Otherwise, it is a substantive behavioural assumption. In effect it says: even if agents in $\mathcal{N}^t$ disagree with the majority decision at time t, their opinions are not strong enough to motivate them to disclose any further evidence. Let $\mathcal{I}^* := \mathcal{I} \setminus \mathcal{N}^0$ be the set of agents in $\mathcal{I}$ who are *not* internally neutral. Here is our second result.

Theorem 2 *In the parallel protocol, there is a unique deliberative equilibrium. If Assumptions (A), (B) and (C) hold, $E^0 = 0$, and $|\mathcal{I}^*| \leq 4$, then this equilibrium is full-disclosure equivalent.*

Scenarios similar to Examples (b)-(d) in Section 3.1 show why Assumptions (A) and (B) and the hypothesis $E^0 = 0$ are needed for the conclusion of Theorem 2. Assumption (C) and the condition $|\mathcal{I}^*| \leq 4$ are necessary because otherwise the balance of public evidence can lurch to an extreme negative or positive value, causing the group to get "stuck" in a suboptimal equilibrium, as shown in the next two examples.

Examples. (e) To see why Assumption (C) is needed, let $\mathcal{I} = \{j, \ell, m, n\}$, and suppose that $\mathcal{C}^0 = \emptyset$ and $\mathcal{M}_j = \{\oplus_j\}$, while $\mathcal{M}_i = \{\oplus_i, \oslash_i\}$ for each $i \in \{\ell, m, n\}$, where $\oplus_j, \oplus_\ell, \oplus_m, \oplus_n$ are positive signals and $\oslash_\ell, \oslash_m, \oslash_n$ are negative. Thus, $\mathcal{N}^0 = \{\ell, m, n\}$, and Assumptions (A) and (B) are satisfied. At the beginning of deliberation, $s_\ell^0 = s_m^0 = s_n^0 = 0$, but $\text{Maj}^0 = 1$ because $s_j^0 = 1$. Since ℓ, m and n are neutral, they dissent from this majority opinion. Suppose that, in contradiction to Assumption (C), they all disclose negative evidence in Round 0. Thus, $\mathcal{C}^1 = \{\oslash_\ell, \oslash_m, \oslash_n\}$, $\mathcal{P}_j^1 = \{\oplus_j\}$, and $\mathcal{P}_i^1 = \{\oplus_i\}$ for each $i \in \{\ell, m, n\}$. Thus, $s_j^1 = s_\ell^1 = s_m^1 = s_n^1 = -1$, so that $\text{Maj}^1 = -1$, and a unanimous deliberative equilibrium is reached in Round 1. But this equilibrium is clearly *not* full-disclosure equivalent, because the group has four pieces of positive evidence and only three pieces of negative evidence, so the full-disclosure majority decision would have been positive.

(f) To see why the condition $|\mathcal{I}^*| \leq 4$ is needed, suppose $\mathcal{I} = \mathcal{I}^+ \sqcup \mathcal{I}^- \sqcup \mathcal{N}$, where $\mathcal{N}$ is the set of internally neutral agents, while agents in $\mathcal{I}^+$ (resp. $\mathcal{I}^-$) initially have positive (resp. negative) opinions. Suppose $|\mathcal{I}^+| = 3$, $|\mathcal{I}^-| = 2$ (so that $|\mathcal{I}^*| = 5$) and $|\mathcal{N}| = 2$ (so Assumption (B) is satisfied). Suppose that $\mathcal{C}^0 = \emptyset$, and for all three $i \in \mathcal{I}^+$, suppose $|\mathcal{M}_i| = 2$ and $y_m = +1$ for both $m \in \mathcal{M}_i$. Meanwhile, for both $i \in \mathcal{I}^-$, suppose $|\mathcal{M}_i| = 3$ and $y_m = -1$ for all three $m \in \mathcal{M}_i$. Finally, suppose the sets $\mathcal{M}_i$ (for $i \in \mathcal{I}$) are all disjoint (so Assumption (A) is also satisfied). Thus,

$$\sum_{i \in \mathcal{I}} \sum_{m \in \mathcal{M}_i} y_m \quad = \quad \sum_{i \in \mathcal{I}^+} \sum_{m \in \mathcal{M}_i} y_m + \sum_{i \in \mathcal{I}^-} \sum_{m \in \mathcal{M}_i} y_m \quad = \quad 3 \times 2 + 2 \times (-3) \quad = \quad 0,$$

so that the consensus opinion after full disclosure would be neutral.

Now, $|\mathcal{I}^+| = 3 > 2 = |\mathcal{I}^-|$, so $\text{Maj}^0 = 1$. In the parallel protocol, *both* agents in $\mathcal{I}^-$ dissent in period 0, and each discloses one piece of negative evidence. So at the start of period 1, $\sum_{c \in \mathcal{C}^1} y_c = -2$. Thus, $s_n^1 = -1$ for both $n \in \mathcal{N}$. At this point, we have

$$\sum_{p \in \mathcal{P}_1^i} y_p + \sum_{c \in \mathcal{C}^1} y_c \quad = \quad 2 - 2 \quad = \quad 0 \quad \text{so that} \quad s_i^1 \quad = \quad 0 \quad \text{for all three } i \in \mathcal{I}^+, \text{ while}$$

$$\sum_{p \in \mathcal{P}_1^i} y_p + \sum_{c \in \mathcal{C}^1} y_c \quad = \quad -2 - 2 \quad = \quad -4 \quad \text{so that} \quad s_i^1 \quad = \quad -1 \quad \text{for both } i \in \mathcal{I}^-.$$

Thus, all agents in $\mathcal{I}^+$ have neutral opinions at the start of period 1, but both agents in $\mathcal{I}^-$ retain their negative opinions. Meanwhile, $s_n^1 = -1$ for both $n \in \mathcal{N}$, so that $\text{Maj}^1 = -1$. Thus, during period 1, the three (now neutral) agents in $\mathcal{I}^+$ dissent, and each discloses one piece of positive evidence, so that $\sum_{c \in \mathcal{C}^2} y_c = -2 + 3 = 1$. Thus, $s_n^2 = 1$, for both $n \in \mathcal{N}$. At this point, we have

$$\sum_{p \in \mathcal{P}_2^i} y_p + \sum_{c \in \mathcal{C}^2} y_c \quad = \quad 1 + 1 \quad = \quad 2 \quad \text{so that} \quad s_i^2 \quad = \quad 1 \quad \text{for all three } i \in \mathcal{I}^+, \text{ while}$$

$$\sum_{p \in \mathcal{P}_2^i} y_p + \sum_{c \in \mathcal{C}^2} y_c \quad = \quad -2 + 1 \quad = \quad -1 \quad \text{so that} \quad s_i^2 \quad = \quad -1 \quad \text{for both } i \in \mathcal{I}^-.$$

Thus, all agents in $\mathcal{I}^*$ have the same opinions at the start of round 2 as in round 0. But $s_n^2 = 1$ for both $n \in \mathcal{N}$ so that $\text{Maj}^2 = 1$. Thus, the two negative agents again dissent, so they disclose information during round 2. Thus, $\sum_{c \in \mathcal{C}^3} y_c = 1 - 2 = -1$, so that $s_n^3 = -1$ for both $n \in \mathcal{N}$. Meanwhile,

$$\sum_{p \in \mathcal{P}_3^i} y_p + \sum_{c \in \mathcal{C}^3} y_c \quad = \quad 1 - 1 \quad = \quad 0 \quad \text{so that} \quad s_i^3 \quad = \quad 0 \quad \text{for all three } i \in \mathcal{I}^+, \text{ while}$$

$$\sum_{p \in \mathcal{P}_3^i} y_p + \sum_{c \in \mathcal{C}^3} y_c \quad = \quad -1 - 1 \quad = \quad -2 \quad \text{so that} \quad s_i^3 \quad = \quad -1 \quad \text{for both } i \in \mathcal{I}^-.$$

So the two agents in $\mathcal{I}^-$ remain negative at the start of round 3, but the three agents in $\mathcal{I}^+$ are now neutral, so that $\text{Maj}^3 = -1$. Thus, the three (now neutral) agents in $\mathcal{I}^+$ dissent, so they disclose their remaining positive evidence during round 3. Thus, $\sum_{c\in\mathcal{C}^4} y_c = -1+3 = 2$, so that $s_n^4 = 1$ for both $n \in \mathcal{N}$. Meanwhile,

$$\sum_{p\in\mathcal{P}_4^i} y_p + \sum_{c\in\mathcal{C}^4} y_c \;=\; 0+2 \;=\; 2 \quad \text{so that} \quad s_i^4 \;=\; 1 \quad \text{for all three } i \in \mathcal{I}^+, \text{ while}$$
$$\sum_{p\in\mathcal{P}_4^i} y_p + \sum_{c\in\mathcal{C}^4} y_c \;=\; -1+2 \;=\; 1 \quad \text{so that} \quad s_i^4 \;=\; 1 \quad \text{for both } i \in \mathcal{I}^-.$$

Thus, at the end of round 4, *all* agents have positive opinions, so deliberative equilibrium has been reached with a positive consensus opinion. This is *not* the same as the neutral consensus opinion which would have been reached with full disclosure. Thus, the deliberative equilibrium is *not* full-disclosure equivalent. ♢

Remark. Note that the conditions of Theorems 1 and 2 are sufficient but not *necessary* for full-disclosure equivalence. For example, if all agents begin with the same information (i.e. $\mathcal{M}_i = \mathcal{M}$ for all $i \in \mathcal{I}$), then the group immediately reaches a full-disclosure equivalent deliberative equilibrium, whether or not Assumptions (A), (B) or (C) or the other hypotheses of Theorems 1 and 2 are satisfied. We do not yet know of conditions which are both necessary and sufficient for full-disclosure equivalence.

3.3 Advisory councils

So far, we have considered deliberation among informed agents who themselves will vote on the final decision. But in many deliberative situations, the informed agents cannot vote, but can only *advise* a decision-maker. This is the case, for example, in "advisory councils" of scientific experts. For brevity, we will refer to the informed agents as *advisors*, and refer to the decision-maker as the *President*. We will assume that the President is *internally neutral* as in Assumption (B), and she acts in "good faith" —she just wants to make the correct decision.[12] Thus, her decision D^t at time t is entirely determined by the balance of publicly disclosed evidence at time t:

$$D^t \quad = \quad \text{sign}\left(\sum_{c\in\mathcal{C}^t} y_c\right). \tag{3D}$$

Each advisor will disclose evidence to sway the President's opinion towards whatever *she* currently believes to be the correct answer. As before, we assume that advisors only disclose evidence when they *dissent* from the President's current decision. However, as evidence is revealed, the advisors themselves may change their opinions. Again, we can

[12]The decision-maker could also be a group (e.g. a legislative body). In this case, we assume all members of this group satisfy these assumptions.

consider two deliberation protocols; in the *serial advisory protocol*, only one new piece of evidence is disclosed during each round of deliberation, whereas in the *parallel advisory protocol*, all dissenting advisors can disclose evidence during each round. The group reaches *deliberative equilibrium* when every advisor stops disclosing evidence —either because she agrees with the President's current opinion, or because she has no countervailing evidence left to disclose. In contrast to Examples (a) and (c), such a deliberative equilibrium must *always* involve unanimous consensus. To see this, suppose by contradiction that advisor i disagreed with the President at time t —say, $D^t = 1$ while $s_i^t \leq 0$. Comparing formulae (3B) and (3D), we deduce that $\sum_{p \in \mathcal{P}_i^t} y_p < 0$ —in other words, i has undisclosed negative evidence. But in this case, she would disclose it, which means we could not yet be in equilibrium.[13]

A deliberative equilibrium is *full-disclosure equivalent* if the President's decision (3D) is the same as it would have been if *all* private evidence had been disclosed. Versions of Theorems 1 and 2 continue to be true in this setting. But the presence of an external, neutral President means that we can eliminate Assumptions (B) and (C).

Theorem 3 *Assume (A), and suppose that either $E^0 = 0$, or there is initial disagreement. Then the serial advisory protocol always reaches a deliberative equilibrium that is full-disclosure equivalent.*

Theorem 4 *Assume (A), and suppose $E^0 = 0$, and $|\mathcal{I}^*| \leq 4$. Then the parallel advisory protocol always reaches a deliberative equilibrium that is full-disclosure equivalent.*

Scenarios similar to Example (b), (d) and (f) in Sections 3.1 and 3.2 show why the remaining hypotheses of Theorems 3 and 4 are still required.

4 Deliberating probabilistic beliefs

In Section 3, we assumed that there were only two possible states of the world (i.e. $|\mathcal{X}| = 2$). We will now suppose there are *many* possible states of the world (i.e. $\mathcal{X}$ is any finite set). Recall from Section 2 that $\mathcal{Y}$ is the set of possible signal-values. Unlike Section 3, we now allow $\mathcal{Y}$ to have any size (even infinite). For all $i \in \mathcal{I}$, $\mathcal{M}_i$ is the set of independent $\mathcal{Y}$-valued random variables ("evidence") observed by i. For all $t \in \mathbb{N}$, recall that agent i's probabilistic beliefs at time t are given by the function B_i^t defined in formula (2E), where π_i is i's prior probability distribution, $\mathcal{P}_i^t$ is her undisclosed evidence, and $\mathcal{C}^t$ is the set of publicly available evidence at time t.

In Section 3, we assumed that each agent could only communicate her binary opinion (as defined by formula (3A)) to other agents, and the group used majority vote (3C) to aggregate these opinions. But if $|\mathcal{X}| \geq 3$ then there is no canonical way to convert a probabilistic belief B over $\mathcal{X}$ into an $\mathcal{X}$-valued "opinion" analogous to formula (3A).[14]

[13] The internal neutrality of the President is crucial for this argument.

[14] If $\mathcal{X}$ was an ordered set or a metric space, then we could map B to its *median point*. If $\mathcal{X}$ was a subset of $\mathbb{R}^N$, then we could map B to the point closest to its *mean*. If $\mathcal{X}$ was just an abstract set, then we could map B to its *mode*, as in formula (3A). In general, these three values will disagree.

Therefore, we will no longer suppose that the group tries to aggregate the *opinions* of its members about $\mathcal{X}$ —instead, the group directly aggregates their *probabilistic beliefs*.

Formally, at each time t, we suppose the agents aggregate their provisional beliefs $(B_i^t)_{i\in\mathcal{I}}$ using a probability aggregation rule $F : \Delta(\mathcal{X})^{\mathcal{I}} \longrightarrow \Delta(\mathcal{X})$, so that the provisional collective belief at time t is given by $F\left((B_i^t)_{i\in\mathcal{I}}\right)$. Assume that F satisfies the ***unanimity*** property that $F(B, B, \cdots, B) = B$ for any belief $B \in \Delta(\mathcal{X})$.[15] We say that ***deliberative equilibrium*** is obtained at time T if all agents agree with the collective belief —i.e. $B_j^T = F\left((B_i^T)_{i\in\mathcal{I}}\right)$ for all $j \in \mathcal{I}$. It is easy to see that such an equilibrium occurs if and only if all agents have the *same* beliefs at time T —that is, if $B_i^T = B_j^T$ for all $i, j \in \mathcal{I}$.

One question is whether such equilibria even exist. Another is whether they have desirable epistemic properties. We will answer these questions shortly.

We will suppose that agents disclose evidence according to some protocol. But unlike Section 3, we will not model this protocol in detail. For any $x \in \mathcal{X}$, let

$$\alpha(x) \quad := \quad \prod_{m\in\mathcal{M}} \rho_x(y_m). \tag{4A}$$

This is the likelihood function that would be used for Bayesian updating by someone who knew *all* the evidence. We will say that ***full information pooling*** occurs at time t if

$$\prod_{c\in\mathcal{C}^t} \rho_x(y_c) \quad = \quad \alpha(x), \qquad \text{for all } x \in \mathcal{X}. \tag{4B}$$

In other words, given the evidence which is publicly available at time t, an agent who starts with no private information will reach the same beliefs that she would have reached if *all* evidence had been disclosed. Note that this does *not* mean that $\mathcal{C}^t = \mathcal{M}$. Clearly, the equality $\mathcal{C}^t = \mathcal{M}$ would be sufficient to obtain (4B), but it is not necessary; in practice, a small but suitably selected subset of $\mathcal{M}$ could achieve full information pooling. We say the agents have a ***common prior*** if $\pi_i = \pi_j$ for all $i, j \in \mathcal{I}$.

Proposition 4.1 *If the agents have a common prior, then there always exists a deliberative equilibrium with full information pooling.*

The proof of Proposition 4.1 is quite simple: suppose that every agent reveals *all* of her evidence. Then there is full information pooling, and it can be checked that all agents end up with the same beliefs; hence they are in equilibrium. (See Appendix B for details.)

Thus, if agents have a common prior, then there is at least one "good" equilibrium. But this equilibrium is not very interesting, since it supposes that the agents have revealed *all* their information, which seems unlikely or even impossible in many real deliberation contexts. Furthermore, there could also be many other deliberative equilibria where there is *not* full information pooling. Indeed, it might not even be possible to reach the equilibrium described in Proposition 4.1, because every possible "deliberation path" might get stuck

[15] For example, linear pooling rules and geometric pooling rules both satisfy this axiom (Genest and Zidek, 1986; Dietrich and List, 2016).

in one of these bad equilibria before it can get to the good equilibrium in Proposition 4.1. So we will now turn our attention to the other equilibria.

We will focus on deliberative groups that have one or more internally neutral agents. We will say that an agent n is *internally neutral* if there is some constant $C > 0$ such that

$$\prod_{m \in \mathcal{M}_n} \rho_x(y_m) \quad = \quad C, \qquad \text{for all } x \in \mathcal{X}. \tag{4C}$$

In other words, n's initial private information does not predispose her to believe any state is any more likely than any other state. For example, this would be the case if $\mathcal{M}_n = \emptyset$ —i.e. n started with no private information at all. We will use the following generalization of Assumption (B) from Section 3:

(B′) There is some agent $n \in \mathcal{I}$ who is internally neutral as in formula (4C).

(Assumption (B) is the special case of Assumption (B′) when $\mathcal{X} = \mathcal{Y} = \{\pm 1\}$ with $\rho(x|x) = p > \frac{1}{2}$ and $\rho(-x|x) = 1 - p$, for both $x \in \mathcal{X}$.) If n is internally neutral, with prior beliefs given by the probability measure $\nu \in \Delta^*(\mathcal{X})$, and B_n^t is her beliefs at time t as in formula (2E), then B_n^t is entirely determined by the publicly available evidence at time t:

$$B_n^t(x) \quad = \quad \frac{\nu(x)}{(\text{SNC})} \prod_{c \in \mathcal{C}^t} \rho_x(y_c), \qquad \text{for all } x \in \mathcal{X}. \tag{4D}$$

In Section 3, just before Subsection 3.1, we introduced Assumption (A), which posited some time t such that for all $i \neq j$, any evidence in $\mathcal{M}_i \cap \mathcal{M}_j$ has been disclosed at or before time t. (Evidence known by only one agent may remain undisclosed.) In particular, if the evidence sets $\{\mathcal{M}_i\}_{i \in \mathcal{I}}$ are disjoint, then this condition is trivially satisfied. Here is the main result of this section.

Theorem 5 *If all agents start with a common prior ν, then a deliberative equilibrium exists. If the group satisfies Assumption (B′), and the equilibrium satisfies Assumption (A), then there is full information pooling, and all agents have beliefs given by $B(x) = \alpha(x)\,\nu(x)/(\text{SNC})$ for all $x \in \mathcal{X}$, where α is as in formula (4A).*

In this result, the *existence* of an equilibrium follows from Proposition 4.1. Theorem 5 does not say that this equilibrium is *unique* —it just says that in *any* such equilibrium, the configuration $\mathcal{C}^t$ of publicly revealed evidence satisfies the full information pooling condition (4B), and thus yields a "maximally informed" decision. Also, Theorem 5 does not tell us the path the group takes to reach equilibrium. We have not modelled the way the deliberation unfolds, the order in which people speak, or what evidence they disclose.

Theorem 5 might not seem too surprising in light of Aumann's (1976) Agreement Theorem. But the models are quite different. Aumann supposes that each agent obtains information through her own personal partition of the state space. Her information is private knowledge, but her information-*partition* is common knowledge. So other agents do not know what she knows, but they know what she *can* know, in principle. They

do not directly communicate their private information, but they can deduce one another's information by observing the discrepancies between their posterior beliefs. Hence "common knowledge of disagreement" is logically impossible between Bayesian agents in his model. In contrast, in our model, agents do *not* know one another's information partitions. But they are able to selectively reveal parts of their private information through communication.

Theorem 5 is actually a special case of a more general result, which does *not* assume a common prior. Suppose the agents have arbitrary prior beliefs $\{\pi_i\}_{i\in\mathcal{I}}$ in $\Delta^*(\mathcal{X})$, and among them there is an internally neutral agent with prior ν. For all other $i \in \mathcal{I}$, define the function $\frac{\mathrm{d}\pi_i}{\mathrm{d}\nu} : \mathcal{X} \longrightarrow \mathbb{R}_+$ by

$$\frac{\mathrm{d}\pi_i}{\mathrm{d}\nu}(x) \quad := \quad \frac{\pi_i(x)}{\nu(x)}, \qquad \text{for all } x \in \mathcal{X}.$$

This measures the deviation between π_i and ν.[16] In particular, if $\pi_i = \nu$, then $\frac{\mathrm{d}\pi_i}{\mathrm{d}\nu} = 1$. If ν is the *uniform measure* (i.e. $\nu(x) = 1/|\mathcal{X}|$ for all $x \in \mathcal{X}$) then $\frac{\mathrm{d}\pi_i}{\mathrm{d}\nu} = |\mathcal{X}| \cdot \pi_i$ for all $i \in \mathcal{I}$.

Let μ be the result of applying the *multiplicative pooling rule*[17] to the beliefs $\{\pi_i\}_{i\in\mathcal{I}}$, with respect to the reference measure ν. That is, for all $x \in \mathcal{X}$,

$$\mu(x) \quad := \quad \frac{\nu(x)}{(\mathrm{SNC})} \prod_{i\in\mathcal{I}} \frac{\mathrm{d}\pi_i}{\mathrm{d}\nu}(x). \tag{4E}$$

Two special cases are of interest. First, if $\pi_i = \nu$ for all $i \in \mathcal{I}$ (as in Theorem 5), then $\mu = \nu$ also. Second, if ν is the uniform measure, then $\mu(x) = \frac{1}{(\mathrm{SNC})} \prod_{i\in\mathcal{I}} \pi_i(x)$ for all $x \in \mathcal{X}$.

Theorem 6 *Suppose the group satisfies Assumption (B′). In any deliberative equilibrium satisfying Assumption (A), all agents have beliefs given by* $B(x) = \alpha(x)\,\mu(x)/(\mathrm{SNC})$ *for all* $x \in \mathcal{X}$*, where* α *and* μ *are as in formulae (4A) and (4E).*

In other words, in deliberative equilibrium, the consensus belief is as if the agents first generated a "synthetic prior" μ via multiplicative pooling (4E), and then shared *all* of their private evidence. In particular, if ν is the uniform measure, then the consensus belief is

$$B(x) \;=\; \frac{1}{(\mathrm{SNC})} \prod_{m\in\mathcal{M}} \rho_x(y_m) \cdot \prod_{i\in\mathcal{I}} \pi_i(x), \qquad \text{for all } x \in \mathcal{X}.$$

Alternately, if $\pi_i = \nu$ for all $i \in \mathcal{I}$, then Theorem 6 yields the outcome of Theorem 5.

Note that Theorem 6 does not say that a deliberative equilibrium exists. In general, there might not be any equilibrium. To see why, recall that in any deliberative equilibrium, all agents must agree. But suppose that the priors $\{\pi_i\}_{i\in\mathcal{I}}$ are quite far apart —in other words, some agents have quite extreme pre-existing biases in their beliefs ("prejudices"), which are not justified by their private information. These agents will try to move other

[16] In fact, $\frac{\mathrm{d}\pi_i}{\mathrm{d}\nu}$ is the *Radon-Nikodym derivative* of π_i with respect to ν. But this is not important here.
[17] See Dietrich (2010, 2019) and Dietrich and List (2016).

agents closer to their point of view by selectively disclosing evidence that "supports" their prejudices, but they might simply run out of such evidence before agreement is achieved. In this case, deliberative equilibrium is obviously impossible.

Even if $\{\pi_i\}_{i\in\mathcal{I}}$ are quite close, an equilibrium might not exist. The problem is that the information in $\{y_m\}_{m\in\mathcal{M}}$ comes in discrete "chunks", so it changes the agents' beliefs by discrete amounts. There might be no way to choose a subset of this evidence that exactly cancels the difference between $\{\pi_i\}_{i\in\mathcal{I}}$. In this case, the best we can hope for is an "approximate" equilibrium, where agents minimize the distance between their beliefs.

We could formally define a general notion of deliberative equilibrium in terms of such distance-minimization —in effect, it would be a pure-strategy Nash equilibrium where each agent's utility function is a decreasing function of the distance between her beliefs and the collective beliefs defined by F. But we will not pursue this for two reasons. First, it is not clear what is the right notion of distance to use. Second, in the resulting game, it would be nontrivial to prove the uniqueness —or even existence —of pure strategy Nash equilibria. So we will leave this as a topic for future research.

A weakness of Theorems 5 and 6 is their need for an internally neutral agent. What if we relax this requirement?

Proposition 4.2 *Suppose that all agents start with a common prior* π (not necessarily uniform)*. Consider any deliberative equilibrium satisfying Assumption* (A)*. Then there is a probability distribution* η (not necessarily π) *such that, in this deliberative equilibrium, all agents have the beliefs* $\alpha \cdot \eta/(\text{SNC})$*, where* α *is as in formula* (4A)*.*

This says that the beliefs in any deliberative equilibrium will be *as if* full information pooling occurred, but all agents began with a common prior η, which is not necessarily the same as π. This does not contradict Proposition 4.1, because that result simply describes a *particular* equilibrium in which $\eta = \pi$. Unfortunately, there might also be *other* deliberative equilibria where $\eta \neq \pi$; indeed, each equilibrium could have a different choice of η. Furthermore these η do not have any obvious normatively appealing interpretation.

5 Related literature

To our knowledge, Klevorick et al. (1984) was the first paper to propose a mathematical model of jury deliberations as an enhancement of the Condorcet Jury Theorem. But the paper by Klevorick et al. is atypical in two ways: first, unlike most of the later literature, it neglects strategic considerations and simply assumes that full information revelation will occur; second, it assumes that the information of each voter takes the form of a continuous (normally distributed) random variable.

Most other models of deliberative democratic decision-making involve discrete information, and emphasize how strategic behaviour can interfere with full information disclosure. For example, in response to earlier results by Austen-Smith and Banks (1996) and Feddersen and Pesendorfer (1996, 1997, 1998, 1999) concerning strategic voting in the setting of the Condorcet Jury Theorem, Coughlan (2000) proposed a cheap-talk model of *strategic*

deliberation between jurors prior to voting. He showed that, as long as all jurors have similar preferences, they will reveal their private information during deliberation. But if their preferences are too heterogeneous, then they may still lie. Since then, several other papers have modelled strategic deliberation as cheap talk (Doraszelski et al., 2003; Austen-Smith and Feddersen, 2006; Meirowitz, 2007; Gerardi and Yariv, 2007; Hummel, 2012; Le Quement, 2013; Le Quement and Yokeeswaran, 2015; Deimen et al., 2015; Rivas and Rodríguez-Álvarez, 2017) or cheap talk mixed with verifiable information (Mathis, 2011). See Landa and Meirowitz (2009) and List (2018,§5.3) for reviews.

Schulte (2010) and Hagenbach et al. (2014) depart from the cheap talk setting, and instead assume that agents cannot lie, but can conceal evidence; this is similar to our model.[18] Schulte (2010) examines a setting where agents may withhold information. But each agent can deduce any information withheld by other agents, because they cannot lie and they are facing a binary state. So when an agent does not reveal her information, this shows that she has information she does not want to reveal, which inadvertently reveals the information itself. Our model has little overlap with Schulte's, because our individuals *can* credibly withhold information.

Hagenbach et al. (2014, §6.3) study an epistemic voting game, in which each voter receives a noisy private signal about the true state, which she may or may not reveal during the initial deliberation stage. They provide necessary and sufficient conditions for an equilibrium in which all voters reveal their private signals. This is an application of their much more general analysis of sequential equilibria in two-period Bayesian games where players can publicly make certifiable statements about their type during the first period. Unlike our model, the agents in the model of Hagenbach et al. form beliefs about one another's private information. Hagenbach et al. assume that each voter receives exactly *one* private signal, whereas in our model each voter could receive any number of private signals. Furthermore, in the model of Hagenbach et al., other voters *know* when a voter i has not fully revealed her private information, and in this case they all assign probability 1 to a particular private signal for voter i. (This follows from their assumption of *extremal beliefs*, along with the strong belief consistency which is a defining property of sequential equilibria.) From this it follows that under fairly general conditions, all voters reveal their private signals in equilibria.

A key element of our model is the presence of "internally neutral" agents who are influenced by the strategic information disclosure of "informed" agents. Some recent papers consider similar scenarios. For example Schnakenberg (2015, 2017) and Jeong (2019) model lobbying as cheap talk by experts whose preferences differ from those of the voters they seek to influence. Jackson and Tan (2013) considers deliberation with hard evidence. In their model, deliberation is followed by a vote; there is a population of experts who may or may not reveal their private verifiable information, and also a group of uninformed voters who may or may not be convinced by these experts. The model differs from the cheap talk settings in that experts' signals are verifiable and the experts simply choose whether

[18] Earlier, Milgrom and Roberts (1986) and Shin (1994) introduced models where agents can either provide verifiable information to a principal or withhold the information. But these models do not involve deliberation or voting.

or not to reveal their signal. But experts and voters may have different preferences, so the experts may behave strategically. In contrast, in our model, everyone has the same preferences (i.e., to find the truth), but *beliefs* are heterogeneous.

The literature described above assumes all deliberation takes place during a single round of communication. But in some models, the *time structure* of information revelation is important. For example, time plays a role in models of herding and information cascades (Banerjee, 1992; Bikhchandani et al., 1992, 1998) and rational learning in social networks (Gale and Kariv, 2003; Acemoğlu et al., 2011; Mueller-Frank, 2013; Mossel et al., 2014b; Lobel and Sadler, 2015, 2016). Imbert et al. (2019) introduce a dynamic deliberation model where agents may lie because of conformism. In another example, Ottaviani and Sørensen (2001, 2006) propose models where heterogeneous experts reveal information, but the order of speech is strategically significant. (In the first paper, experts receive a noisy binary signal about a binary state of the world; in the second, they receive a noisy continuous signal about a continuous state of the world.) The experts want to preserve their reputations by not making wrong forecasts; this can lead to a herding phenomena similar to information cascades.[19] Unlike Ottaviani and Sørensen, we do not consider reputational incentives.

Geanakoplos and Polemarchakis (1982) and Cave (1983) described "communication protocols" by which two Bayesian agents with initially heterogeneous beliefs could reach the agreement of beliefs predicted by Aumann (1976); furthermore, this consensus would be common knowledge. This result was generalized by Bacharach (1985) to any finite set of agents, assuming all-to-all communication. Parikh and Krasucki (1990) and Krasucki (1996) showed that these results partially generalize even without all-to-all communication: given a network of agents linked by unidirectional communication channels, their beliefs can converge to a consensus (assuming sufficient network connectivity), but not common knowledge. This was further extended by Houy and Ménager (2008), who considered the influence of speech order by comparing different communications protocols.

Importantly, in the models of Ottaviani, Sørensen, Parikh, Krasucki, Houy and Ménager, the order of speech is exogenous, while in Imbert et al. (2019), it is random. But in the models in Section 3 of this paper, speech order is endogenous.

As the Aumann-inspired literature shows, not all deliberation models involve strategic dissimulation. For example, Hafer and Landa (2007) consider agents who lack logical omniscience; deliberation consists of exchanging reasons to induce other voters to recognize certain truths. Unlike us, they propose a (non-Bayesian) game-theoretic model. Sethi and Yildiz (2012) suppose individuals first form beliefs after receiving signals, and then deliberate in a sequence of rounds. In each round, individuals truthfully and simultaneously announce their current *beliefs* to the public (rather than their signals as in our model). They then update their beliefs based on each other's announcements. This continues until an equilibrium at which no further belief revision occurs, somewhat analogous to our notion of *deliberative equilibrium*. The social psychology literature on deliberation also typically assumes agents are honest and do not strategically withhold information (Adamowicz et al., 2005; Kerr and Tindale, 2004; Tindale and Kluwe, 2015). Recent experimental investiga-

[19] Visser and Swank (2007) also present a model of deliberation with reputational incentives, but in their model the disclosure of private information is simultaneous.

tion of deliberation in laboratory settings shows that agents publicly and truthfully reveal their private information at a much higher rate than what is predicted by the models of strategic deliberation described above (Goeree and Yariv, 2011; Le Quement and Marcin, 2019); this is consistent with our assumption that agents deliberate in "good faith". In the experimental design of Dickson et al. (2008), agents exchange *reasons* rather than information (similar to the model of Hafer and Landa (2007)). These attempts at persuasion may backfire, so the optimal strategy is sometimes to remain silent. But experimentally, agents are much more candid than the optimal strategy recommends.[20]

Hartmann and Rafiee Rad (2017) analyzed the anchoring effect in non-strategic deliberation. Deliberation proceeds in a sequence of rounds and during each round, group members speak in a fixed order. More recently, Hartmann and Rafiee Rad (2018) presented a Bayesian model for non-strategic rational deliberation, somewhat different than the one in Section 4 of this paper. Their agents have *partial* beliefs. In the course of deliberation, they cast a vote which is based on their beliefs, then they update their beliefs using Bayesian updating rules and the previous votes of the other group members. They have found the conditions under which deliberation results in a consensus and correctly tracks the true state better than the simple voting procedure of the Condorcet Jury Theorem.

There is also an interesting experimental literature on deliberation. Empirically, information that is held by only one member is often omitted from discussion (Stasser et al., 1989). Thus, groups with diverse members often fail to fully exploit their informational diversity (Stasser and Titus, 1985, 1987). Stasser et al. (1995a) and Stewart and Stasser (1995b) conducted deliberation experiments in which some deliberators were assigned the role of "expert" analogous to the "informed" agents of Section 3. We have already mentioned the papers by Dickson et al. (2008, 2015), Goeree and Yariv (2011) and Le Quement and Marcin (2019). Finally, Van Dijk et al. (2014) present an experimental study of judicial decision-making in small groups, which shows that aggregation and deliberation both improve the reliability of group decisions —especially in difficult borderline cases.

The present paper also has common elements with the literature on *opinion dynamics* in social networks.[21] Like this paper, most of that literature assumes that communication is *nonstrategic*.[22] Much of it also assumes that agents are *Bayesian*; this includes the literatures on Aumann's (1976) Agreement Theorem, on information cascades, and on rational learning cited above, but also includes some papers which do not fit neatly in any of these categories, such as Rosenberg et al. (2009) or Jiménez-Martínez (2015). Other models are "quasi-Bayesian": agents aspire to be Bayesian, but make systematic mistakes – for example, failing to account for correlations between other agents (Bala and Goyal, 1998, 2001; DeMarzo et al., 2003; Neilson and Winter, 2008), or mixing Bayesian and non-Bayesian inference rules (Jadbabaie et al., 2012).

But there are also important differences. We suppose that agents disclose *evidence*,

[20]The experimental design of Dickson et al. (2015) is similar, but it involves two opposed speakers attempting to persuade an audience through public debate, and investigates the consequences of different debate formats on the candour of the debaters.

[21]Castellano et al. (2009), Acemoğlu and Ozdaglar (2011) and Mossel and Tamuz (2017) each survey parts of this literature.

[22]Exceptions include Rosenberg et al. (2009) and Mossel et al. (2015).

whereas in opinion dynamics models they just reveal *beliefs* —either explicitly through communication, or implicitly through their strategic behaviour. In opinion dynamics models, agents only *learn* from their peers, whereas in our model, they also seek to *persuade* their peers. Most opinion dynamics models assume agents communicate through a social network; the topology of this network often has important implications for convergence to a social consensus. In contrast, we assume everyone communicates directly with everyone else. Relatedly, much of the opinion dynamics literature focuses on the asymptotic properties of very large societies, whereas we are interested mainly in small groups.

6 Discussion

Our results depend on several assumptions. First and most obviously, we assume that the agents deliberate in *good faith*: they only wish for the group decision to be correct, and do not seek to promote a particular ideology or other agenda. This contrasts with the literature on *strategic deliberation* reviewed at the start of Section 5. This assumption is obviously not plausible in parliamentary debates, internet chat forums, or other politically polarized discussions. But it may be a good approximation for some juries and scientific committees. As already noted at the end of Section 5, nonstrategic communication is also assumed in most of the literature on opinion dynamics and rational learning in social networks, and most of the social psychology literature on deliberation, and confirmed by the experimental results of Goeree and Yariv (2011) and Le Quement and Marcin (2019).

Second and relatedly, it may seem that we have implicitly assumed that all evidence is *verifiable*: agents cannot lie. But we do not *need* this assumption. Since they deliberate in "good faith", agents have *no incentive* to lie. So we do not need information to be verifiable. Nevertheless, such a verifiability assumption is common in the aforementioned literature, and would be appropriate in the sort of deliberative contexts we have in mind.

Third, we have assumed that agents are *Bayesian*. This assumption is ubiquitous in the literatures on Aumann's Agreement Theorem, information cascades, rational learning and other opinion dynamics models discussed in Section 5. But it might be an overly sophisticated model of human cognition. This has motivated the development of "non-Bayesian" models of opinion dynamics, in which an agent adopts a weighted average of her peers' beliefs (DeGroot, 1974; DeMarzo et al., 2003; Neilson and Winter, 2008; Golub and Jackson, 2010; Acemoğlu et al., 2010; Friedkin and Johnsen, 2011; Acemoğlu et al., 2013; Mueller-Frank, 2014), conforms to the majority opinion of her peers (Mossel et al., 2014a; Tamuz and Tessler, 2015), or computes a maximum likelihood estimator (Mossel and Tamuz, 2010).

Fourth, although our agents are Bayesian, they are "myopic" in the sense that each agent always believes her *current* beliefs are correct. Shouldn't each agent interpret the disagreement of other agents as *prima facie* evidence that there are things she doesn't know? This is the logic of Aumann's (1976) Agreement Theorem and the ensuing literature (reviewed in Section 5). But Aumann's model assumes that each agent has a complete understanding of the epistemic capacities of the other agents: she might not know what they know, but she knows what they *could* know. We do not make this assumption. If

agents i and j disagree, then i does not know if this is because she knows things that j does not know, or j knows things that she doesn't know, or both. As explained after Example (a) in Section 3, i does not even know *how much* j knows —does j have only one piece of undisclosed evidence, or one hundred? The only way to discover is through deliberation.

This myopia may also arise from bounded rationality. For example, Eyster and Rabin (2005,§5) and Elbittar et al. (2016) consider Condorcet Jury models with boundedly rational voters who ignore the informational content implicit in the behaviour of other voters; the former apply their theory of *cursed equilibria*, while the latter introduce *subjective beliefs equilibria*.[23] It is beyond the scope of this paper to microfound the myopic behaviour of our deliberating agents using the equilibrium concepts of Eyster and Rabin (2005) or Elbittar et al. (2016). But this would be an interesting direction for future research.

This myopia also explains another feature of our model. If an agent really wants the group to find the truth, then why does she seek to persuade other agents of her *current* beliefs, when she knows that these beliefs may change, and hence are probably incorrect? Such tendentiousness seems like hubris. The reason is simple: although a rational agent knows her current beliefs might not be correct, her current beliefs are nevertheless her current *best guess* about the beliefs she will have in the future after receiving more information.[24] So the best she can do given her *current* information is to advocate her *current* beliefs.

The Nash equilibrium described in Appendix C provides another justification for these behavioural assumptions. Suppose one agent was actually far-sighted and fully aware of her own fallibility. If she expects *other* agents to exhibit myopia and hubris as in the previous two paragraphs, then the analysis of Appendix C says it is her *best response* to behave in a similar manner, because the outcome will be full-disclosure equivalent.

Why must agents advocate any position at all? Why do they not just exchange information in a neutral and unbiased manner, without taking sides? Better yet, why don't they just dump *all* of their evidence on the table, and let the group sort it out? The reason goes to the heart of our model: clearly and credibly communicating evidence —especially complex evidence —is costly in both time and effort, for both the speaker and the listeners. So agents in our model are constrained in how much information they can communicate, as in the models of Glazer and Rubinstein (2001, 2004). Each agent's beliefs might arise from a very large amount of undisclosed evidence (the accumulation of years of formal education, professional practice, academic research, or even life experience in general), and she can feasibly disclose only a small fraction of it. By what criterion can she choose what to disclose? Her only criterion is her current beliefs.

It is also worth commenting on the striking difference between the results in Section 3 and those in Section 4. Theorems 1 and 3 achieves a positive conclusion (full-disclosure equivalence), but only under rather restrictive hypotheses. Those of Theorem 2 and 4 are even more restrictive. Theorems 5 and 6 are sweepingly general by comparison. Why is

[23] Szembrot (2017) and Demange and Van Der Straeten (2017) have also proposed models of boundedly rational voting behaviour based on cursed equilibria.

[24] This is a simple consistency requirement: if i's current beliefs were X, but she believed that in the future she would have beliefs Y, then rationally she should change her beliefs from X to Y immediately.

this? The difference arises from two differing notions of "deliberative equilibrium", which in turn reflect two different interpretations of "good faith" deliberation. In Section 3, agents do not care whether other agents have the correct *beliefs*, as long as the majority vote is correct. Each agent only cares that a majority agrees with her (binary) opinion. So deliberative equilibrium is relatively easy to achieve. But by the same token, such an equilibrium might fail to be full-disclosure equivalent; as the counterexamples in Section 3 demonstrate, it could involve double-counting of evidence or other forms of "groupthink".

In contrast, in Section 4, it is not enough for an agent that the group makes the correct *decision* —she is not satisfied until all other agents *exactly share her* (probabilistic) *beliefs*. This makes deliberative equilibrium harder to achieve, but it ensures that in any such equilibrium, there is full information pooling. Since the agents set a more exacting epistemic standard, they are less likely to fall into groupthink.

This also illustrates the importance of the collective decision rule. In Section 3, the decision rule was a majority vote. But in Section 4, the group seeks to aggregate their probabilistic beliefs. This changes the incentives of each agent, and thus changes the dynamics of the deliberation. To see another example, suppose that instead of a majority in Section 3, the group decision required *unanimity* (as in legal juries). It might seem that this should make no difference, since deliberative equilibria result in unanimity anyways. But the difference is significant. In Section 3, only "dissenting" agents (those in the minority) had an incentive to disclose information; agents on the majority side remained silent. But under a unanimity rule, *all* agents would have an incentive to disclose information, until unanimous consensus is achieved. Paradoxically, this gives more deliberative influence to agents currently in the majority, so that majorities tend to grow larger over time, until they become unanimous. The outcome is often *not* full-disclosure equivalent; Theorems 1 and 2 do *not* hold under a unanimous decision rule.

There are many avenues for future exploration. People are neither sophisticated Bayesians nor simple-minded averagers or conformists. A more realistic model of deliberation requires a more realistic model of human cognition —both *epistemic* cognition (about the facts themselves) and *social* cognition (about other people). Even scientists who aspire to perfect rationality and objectivity and who claim to deliberate in "good faith" have biases and prejudices, and have careers (or egos) invested in particular theories. Perhaps this can be partly captured by relaxing the common-priors assumption made in most of our results, but perhaps it will require a more radical departure.

In important part of deliberation is not the exchange of information, but the exchange of *understanding*. Humans are not logically omniscient, and they often fail to see the implications of the evidence in front of them. Deliberation can help them overcome this limitation (Hafer and Landa, 2007; Maciejovsky and Budescu, 2007). A realistic model of deliberation would also incorporate this phenomenon.

Appendices

A Proofs from Section 3

Proof of the Proposition. The Condorcet Jury Theorem says that the most reliable decision rule is obtained by applying majority rule directly to the signals $\{y_m\}_{m\in\mathcal{M}}$. The outcome is $\text{sign}\left(\sum_{m\in\mathcal{M}} y_m\right)$. For all $i \in \mathcal{I}$, if $\mathcal{M}_i = \mathcal{M}$ then formula (3B) says that $s_i^t = \text{sign}\left(\sum_{m\in\mathcal{M}} y_m\right)$, so that i's vote already matches the most reliable group decision. If this holds for all $i \in \mathcal{I}$, then the majority decision will be the optimal decision. □

All four theorems in Section 3 invoke Assumption (A): there is some $t_0 \in \mathbb{N}$ such that $\mathcal{P}_i^{t_0} \cap \mathcal{P}_j^{t_0} = \emptyset$ for all distinct $i, j \in \mathcal{I}$. For notational simplicity, we will assume throughout Appendix A that $t_0 = 0$ —in other words, we will assume that

$$\mathcal{M}_i \cap \mathcal{M}_j \quad = \quad \emptyset, \qquad \text{for all distinct } i, j \in \mathcal{I}. \tag{A1}$$

This is without loss of generality, because we can "reset the clock" at the first t such that $\mathcal{P}_i^t \cap \mathcal{P}_j^t = \emptyset$ for all distinct $i, j \in \mathcal{I}$, to ensure that (A1) is satisfied. For all $t \in \mathbb{N}$, let

$$E^t \quad := \quad \sum_{c\in\mathcal{C}^t} y_c$$

be the balance of publicly available evidence at time t. For all $i \in \mathcal{I}$, let

$$P_i^t \quad := \quad \sum_{p\in\mathcal{P}_i^t} y_p$$

be the balance of i's undisclosed private evidence at time t. The next lemma will be used repeatedly in the proofs of Theorems 1 to 4.

Lemma A1 *Assume statement* (A1) *is satisfied. Then*

(a) $\displaystyle\sum_{m\in\mathcal{M}\setminus\mathcal{C}^0} y_m \;=\; \sum_{i\in\mathcal{I}} \sum_{m\in\mathcal{M}_i} y_m.$

(b) $\displaystyle\sum_{m\in\mathcal{M}} y_m \;=\; E^t + \sum_{i\in\mathcal{I}} P_i^t$, *for all* $t \in \mathbb{N}$.

Proof: Part (a) follows immediately from statement (A1). Meanwhile, at any time t in deliberation (in either the serial or parallel protocols), statement (A1) implies that

$$E^t + \sum_{i\in\mathcal{I}} P_i^t \quad = \quad \sum_{c\in\mathcal{C}^0} y_c + \sum_{i\in\mathcal{I}} \sum_{m\in\mathcal{M}_i} y_m, \tag{A2}$$

because every time a piece of evidence is added to the common pool $\mathcal{C}^t$, it is removed from the private evidence set of exactly one agent. Combining equation (A2) with part (a), we obtain part (b) of the lemma. □

Proof of Theorem 1. Suppose the group reaches deliberative equilibrium at time t. There are two cases: either $\text{Maj}^t = 0$ or $\text{Maj}^t = \pm 1$.

Case 1. Suppose $\text{Maj}^t = 0$.

Claim 1: *$s_i^t = 0$ for all $i \in \mathcal{I}$.*

Proof: (by contradiction) If $\text{Maj}^t = 0$, then $|\mathcal{I}_+^t| = |\mathcal{I}_-^t|$. Thus, $|\mathcal{I}_+^t| > 0$ if and only if $|\mathcal{I}_-^t| > 0$. Suppose $|\mathcal{I}_+^t| > 0$ and $|\mathcal{I}_-^t| > 0$. There are now two cases:

- If $E^t \leq 0$, then $P_i^t > 0$ for all $i \in \mathcal{I}_+^t$. Thus, voters in $\mathcal{I}_+^t$ have undisclosed positive evidence, and dissent from the (neutral) majority opinion, contradicting the fact that we are in deliberative equilibrium.
- If $E^t \geq 0$, then $P_i^t < 0$ for all $i \in \mathcal{I}_-^t$. Thus, voters in $\mathcal{I}_-^t$ have undisclosed negative evidence, and dissent from the (neutral) majority opinion, contradicting the fact that we are in deliberative equilibrium.

Either way, there is a contradiction. To avoid this, we must have $|\mathcal{I}_+^t| = |\mathcal{I}_-^t| = 0$. ◇ Claim 1

Claim 2: *$E^t = 0$, and $P_i^t = 0$ for all $i \in \mathcal{I}$.*

Proof: (by contradiction) Suppose that $E^t \neq 0$. Then Claim 1 implies that $P_i^t = -E^t$ for all $i \in \mathcal{I}$. This contradicts Assumption (B) (which says there is at least one internally neutral voter). To avoid this contradiction, we must have $E^t = 0$. Then Claim 1 implies that $P_i^t = 0$ for all $i \in \mathcal{I}$. ◇ Claim 2

Since $\text{Maj}^t = 0$, Claim 2 and Lemma A1(b) imply that the equilibrium is full-disclosure equivalent.

Case 2. Suppose $\text{Maj}^t = 1$. (The argument when $\text{Maj}^t = -1$ is very similar.)

Claim 3: *If $E^t > 0$, then for all $i \in \mathcal{I}$, we must have $s_i^t = 1$.*

Proof: (by contradiction) Suppose that $s_i^t \leq 0$ for some $i \in \mathcal{I}$. Then we must have $P_i^t < 0$ (because $E^t > 0$), which means i has undisclosed negative evidence. But since she dissents from the (positive) majority decision, she would disclose this evidence, contradicting the fact that we are in equilibrium. ◇ Claim 3

Claim 4: *$E^t \leq 0$.*

Proof: (by contradiction) Suppose that $E^t > 0$. Each round of deliberation involves disclosure of exactly one piece of evidence, so we must have $E^t = E^{t-1} \pm 1$. Suppose $E^t = E^{t-1} - 1$. Then during round $t-1$ of deliberation, someone disclosed a negative signal —say, agent j. Thus, $s_j^{t-1} = s_j^t$, since agents do not change their own opinion when they disclose information. But $s_j^t = 1$ by Claim 3, so j would have no reason to disclose negative information during round $t-1$. So this is impossible.

It follows that we must have $E^t = E^{t-1} + 1$, which means someone must have disclosed *positive* information during round $t-1$ —again, say it is agent j. The only reason for j to do this would be to dissent from a neutral or negative majority decision. Thus, either $\text{Maj}^{t-1} = 0$ and $s_j^{t-1} = 1$, or $\text{Maj}^{t-1} = -1$ and $s_j^{t-1} = 0$ or 1.

If $\text{Maj}^{t-1} = -1$, then $\mathcal{I}_-^{t-1} \neq \emptyset$. On the other hand, if $\text{Maj}^{t-1} = 0$ and $s_j^{t-1} = 1$, then $\mathcal{I}_+^{t-1} \neq \emptyset$ (because $j \in \mathcal{I}_+^{t-1}$). Thus, we must also have $\mathcal{I}_-^{t-1} \neq \emptyset$ (because otherwise we could not have $\text{Maj}^{t-1} = 0$). Either way, $\mathcal{I}_-^{t-1} \neq \emptyset$.

For all $i \in \mathcal{I}_-^{t-1}$, we must have $P_i^{t-1} \leq -E^{t-1} - 1 = -E^t$. But $P_i^t = P_i^{t-1}$, because i did not disclose any information during round $t-1$ (j did). Thus, $P_i^t \leq -E^t$. But then we would have $s_i^t \leq 0$, contradicting Claim 3. To avoid this contradiction, we must have $E^t \leq 0$. ◇ Claim 4

Claim 5: *If $E^t \leq 0$, then $P_i^t \geq 0$ for all $i \in \mathcal{I}$, and $P_j^t > |E^t|$ for some $j \in \mathcal{I}$.*

Proof: (by contradiction) Suppose $P_i^t < 0$ for some $i \in \mathcal{I}$. Then i has undisclosed negative evidence. Furthermore $s_i^t = -1$ (because $E^t \leq 0$), so i dissents from the (positive) majority decision, so she will disclose some of her negative evidence, contradicting the fact that we are in deliberative equilibrium. This proves the first statement.

To prove the second statement, note that if $\text{Maj}^t = 1$, then we must have $s_j^t = 1$ for at least some $j \in \mathcal{I}$. But if $s_j^t = 1$, then $P_j^t + E^t > 0$, hence $P_j^t > |E^t|$. ◇ Claim 5

Claim 4 says that $E^t \leq 0$. Thus, for all $i \in \mathcal{I}$, Claim 5 implies that $P_i^t \geq 0$, and $P_j^t > |E^t|$ for some $j \in \mathcal{I}$. Thus,

$$\sum_{m \in \mathcal{M}} y_m \underset{(\diamond)}{=} E^t + \sum_{i \in \mathcal{I}} P_i^t = E^t + P_j^t + \sum_{i \in \mathcal{I} \setminus \{j\}} P_i^t \underset{(*)}{>} 0 + \sum_{i \in \mathcal{I} \setminus \{j\}} P_i^t \underset{(\dagger)}{\geq} 0,$$

where $(\diamond)$ is by the Lemma A1(b), $(*)$ is because $P_j^t > |E^t|$, and $(\dagger)$ is because $P_i^t \geq 0$ for all $i \in \mathcal{I}$. Thus, the equilibrium is full-disclosure equivalent, because $\text{Maj}^t = 1$. □

Proof of Theorem 2. The parallel protocol converges to a unique equilibrium because it is deterministic, and must end in finite time because $\mathcal{I}$ is finite and $\mathcal{M}_i$ is finite for all $i \in \mathcal{I}$. Now assume (A), (B), (C), and $E^0 = 0$. As explained at the start of Appendix A, we can assume without loss of generality that Assumption (A) is satisfied in the form of statement (A1), hence invoke Lemma A1. Let $\mathcal{I}^* = \mathcal{I}^+ \sqcup \mathcal{I}^-$, where agents in $\mathcal{I}^+$ (resp. $\mathcal{I}^-$) initially have positive (resp. negative) opinions. If $\mathcal{I}^* = \emptyset$, then deliberative equilibrium is reached immediately with a neutral consensus, which is obviously full-disclosure equivalent. So we will assume that $\mathcal{I}^* \neq \emptyset$; hence $\mathcal{I}^+ \neq \emptyset$ or $\mathcal{I}^- \neq \emptyset$. Let $\mathcal{N}$ be the set of internally neutral agents; thus, $|\mathcal{N}| \geq 1$ by Assumption (B). There are eleven cases to consider.

Case (i). Suppose $|\mathcal{I}^-| = 0$. Then $\text{Maj}^0 = 1$ because $s_i^0 = 1$ for all $i \in \mathcal{I}^+$ and $s_n^0 = 0$ for all $n \in \mathcal{N}$. Agents in $\mathcal{N}$ do not disclose any evidence, by Assumption (C). Thus, the group is immediately in deliberative equilibrium that is full-disclosure equivalent.

Case (ii). If $|\mathcal{I}^+| = 0$, then we have a full-disclosure equivalent equilibrium, by a similar argument to *Case 1*.

Case (iii). Suppose $|\mathcal{I}^+| = |\mathcal{I}^-| = 1$. Let $\mathcal{I}^+ = \{j\}$ and $\mathcal{I}^- = \{k\}$. Let j' be a fictional agent such that $\mathcal{M}_{j'}$ contains no negative evidence and P_j^0 pieces of positive evidence. Let k' be a fictional agent such that $\mathcal{M}_{k'}$ contains no positive evidence and $|P_k^0|$ pieces of negative evidence. For all $n \in \mathcal{N}$, let n' be a fictional agent with $\mathcal{M}_{n'} = \emptyset$. Let $\mathcal{I}' := \{i'; i \in \mathcal{I}\}$. Note that the total balance of evidence for $\mathcal{I}'$ is the same as it was for $\mathcal{I}$, so the full-disclosure decision for the two groups is the same. Deliberation amongst the agents of $\mathcal{I}'$ in the serial protocol is identical to deliberation amongst the agents of $\mathcal{I}$ in the parallel protocol with Assumption (C).[25] Theorem 1 says that $\mathcal{I}'$ reaches a full-disclosure equivalent equilibrium in the serial protocol. Thus, $\mathcal{I}$ reaches a full-disclosure equivalent equilibrium in the parallel protocol with Assumption (C).

Case (iv). Suppose $|\mathcal{I}^+| = 2$, $|\mathcal{I}^-| = 1$, and $|\mathcal{N}| = 1$. Let $\mathcal{N} = \{n\}$, let $\mathcal{I}^- = \{i\}$, let $\mathcal{I}^+ = \{j, k\}$ and suppose that $P_j^0 \leq P_k^0$. Deliberation unfolds as follows.

Round 0. $\text{Maj}^0 = 1$, so i dissents and discloses negative evidence. (Agent n does not disclose anything, by Assumption (C).) Thus, $E^1 = -1$.

Round 1. $E^1 = -1$, so $s_i^1 = s_n^1 = -1$, while $s_j^1 \geq 0$ and $s_k^1 \geq 0$. There are two cases.

- If $s_j^1 = 0$ or $s_k^1 = 0$, then $\text{Maj}^1 = -1$. Both j and k dissent and disclose positive evidence. Thus, $E^2 = E^1 + 2 = 1$ by statement (A1). Go to *Round 2*.
- If $s_j^1 = s_k^1 = 1$, then $\text{Maj}^1 = 0$. Again, j and k dissent and disclose positive evidence. But now i also dissents. If $P_i^1 = 0$, then i has no evidence left to disclose, so $E^2 = E^1 + 2 = 1$ by statement (A1); go to *Round 2*. However, if $P_i^1 \leq -1$, then i discloses negative evidence, so $E^2 = E^1 + 2 - 1 = 0$ by statement (A1). In this case, skip *Round 2* and go to *Round 3*.

Round 2. $E^2 = 1$. Thus, $s_n^2 = s_j^2 = s_k^2 = 1$, so $\text{Maj}^2 = 1$. There are two cases:

- If $P_i^2 = 0$, then $s_i^2 = 1$, so we reach a deliberative equilibrium that is full-disclosure equivalent.
- If $P_i^2 \leq -1$, then i dissents and discloses negative evidence. Thus, $E^3 = E^2 - 1 = 0$, so go to *Round 3*.

Round 3. We now have $E^3 = 0$, just as in *Round 0*. There are five possible scenarios.

- If P_i^3, P_j^3, and P_k^3 are all nonzero, then return to *Round 0* and repeat the argument found there.
- If $P_i^3 = 0$ while P_j^3 or P_k^3 are nonzero, then we have entered the situation described in *Case (i)* above, so apply the argument provided there.
- Likewise, if P_i^3 is nonzero while $P_j^3 = P_k^3 = 0$, then we have entered *Case (ii)*.
- If $P_k^3 = 0$ while P_i^3 and P_j^3 are nonzero (or if $P_j^3 = 0$ while P_i^3 and P_k^3 are nonzero), then we have entered *Case (iii)*.

[25] We replaced each i in $\mathcal{I}$ with the corresponding i' in $\mathcal{I}'$ to replicate the effect of Assumption (C).

- If $P_i^3 = P_i^3 = P_j^3 = 0$, then $\text{Maj}^3 = 0$. This is a full-disclosure equivalent equilibrium.

The process described above eventually terminates when one or more agents runs out of evidence. In all terminal states, the deliberative equilibrium is full-disclosure equivalent.

Case (v). Suppose $|\mathcal{I}^+| = 2$, $|\mathcal{I}^-| = 1$, and $|\mathcal{N}| \geq 2$. Let $\mathcal{I}^- = \{i\}$ and $\mathcal{I}^+ = \{j, k\}$. At any time, the deliberating group can be in one of several "states" (defined below). Deliberation starts in the state *Start*. The rules determining transition from one state to another are described below. This state-transition process eventually terminates in some equilibrium, because the agents eventually deplete their evidence. We must show that all these equilibria are full-disclosure equivalent. Here are the possible states.

Start. $E^0 = 0$, so that $s_n^0 = 0$ for all $n \in \mathcal{N}$. Meanwhile $P_i^0 \leq -1$, $P_j^0 \geq 1$ and $P_k^0 \geq 1$, so that $s_i^0 = -1$ while $s_j^0 = s_k^0 = 1$; thus, $\text{Maj}^0 = 1$. Agent i dissents and discloses negative evidence. (The agents in $\mathcal{N}$ do not disclose anything, by Assumption (C).) Thus, $E^1 = E^0 - 1 = -1$, so go to *State -1*.

State -1. $E^t = -1$, so that $s_n^t = -1$ for all $n \in \mathcal{N}$, and $s_i^t = -1$, so that $\text{Maj}^t = -1$. Meanwhile, $P_j^t \geq 1$ and $P_k^t \geq 1$, so that $s_j^t \geq 0$ and $s_k^t \geq 0$. So j and k dissent and disclose positive evidence. Thus, $E^{t+1} = E^t + 2 = 1$ by statement (A1), so go to *State 1*.

State 1. $E^t = 1$, so that $s_j^t = s_k^t = 1$ and $s_n^t = 1$ for all $n \in \mathcal{N}$. Thus, $\text{Maj}^t = 1$. There are two subcases: either $P_i^t = 0$ or $P_i^t \leq -1$.

- If $P_i^t = 0$, then $s_i^t = 1$, so the group is in deliberative equilibrium. Since $P_j^3 \geq 0$ and $P_k^3 \geq 0$ while $P_i^t = 0$, the majority decision is full-disclosure equivalent.
- If $P_i^t = -1$, then $s_i^t = 0$, while if $P_i^t \leq -2$, then $s_i^t = -1$. Either way, i dissents and discloses negative evidence. Thus, $E^{t+1} = E^t - 1 = 0$, so go to *State 0*.

State 0. $E^t = 0$, so that $s_n^t = 0$ for all $n \in \mathcal{N}$. There are now four subcases.

- If $P_i^t = 0$ then go to *Case (i)*.
- If $P_i^t \leq -1$ and $P_j^t = P_k^t = 0$, then go to *Case (ii)*.
- If $P_i^t \leq -1$, while $P_j^t = 0$ and $P_k^t \geq 1$ (or vice versa), then go to *Case (iii)*.
- If $P_i^t \leq -1$, $P_j^t \geq 1$ and $P_k^t \geq 1$, then go back to the state *Start*.

In all equilibria of this state-transition system, the decision is full-disclosure equivalent.

Case (vi). Suppose $|\mathcal{I}^+| = |\mathcal{I}^-| = 2$. During the initial rounds of deliberation, both agents in $\mathcal{I}^+$ disclose positive evidence, and both agents in $\mathcal{I}^-$ disclose negative evidence, but the balance of public evidence remains neutral. This continues until at least one of the agents in $\mathcal{I}^\pm$ runs out of evidence, and we enter one of *Cases (i)-(v)* above.

Case (vii). Suppose $|\mathcal{I}^+| = 3$, $|\mathcal{I}^-| = 1$, and $|\mathcal{N}| = 1$. Let $\mathcal{I}^- = \{i\}$, let $\mathcal{I}^+ = \{j, k, \ell\}$, and let $\mathcal{N} = \{n\}$. At the start of deliberation, $\text{Maj}^0 = 1$, so i dissents and discloses a

piece of negative evidence. There is then a back-and-forth exchange of evidence between i and the three agents in $\mathcal{I}^+$, until *either* one side persuades the other, *or* at least one of these four agents runs out of private evidence. (Agent n never discloses anything, by Assumption (C).) There are four scenarios, depending on who (if anyone) runs out of private evidence first.

Scenario 1. Suppose that the last agent amongst j, k and ℓ runs out of evidence in the same round as i does. Thus, there is some time t such that $E^t = P_j^0 + P_k^0 + P_\ell^0 + P_i^0$, because all four agents have disclosed all their evidence, and these evidence sets are disjoint by statement (A1). Since no agent has any private evidence remaining, all of their opinions are determined by the public evidence: we have $s_i^t = s_n^t = s_j^t = s_k^t = s_\ell^t = \text{sign}(E^t)$. Thus, the group reaches a unanimous, full-disclosure equivalent deliberative equilibrium at time t.

Scenario 2. Suppose that j, k and ℓ all run out of evidence before i does. Thus, there is some time t such that $E^t = P_j^0 + P_k^0 + P_\ell^0 - Q_i$ for some $Q_i < |P_i^0|$, because j (resp. k, ℓ) has disclosed all of her P_j^0 (resp. P_k^0, P_ℓ^0) pieces of positive evidence, while i has disclosed only Q_i pieces of negative evidence. (Again, these evidence sets are disjoint by statement (A1).) Thus, $P_i^t = P_i^0 + Q_i$, and $P_i^t < 0$. Note that $E^t + P_i^t = (P_j^0 + P_k^0 + P_\ell^0 - Q_i) + (P_i^0 + Q_i) = P_j^0 + P_k^0 + P_\ell^0 + P_i^0$ —that is, the total balance of *all* evidence. Meanwhile, $P_j^t = P_k^t = P_\ell^t = 0$, so $s_j^{t'} = s_k^{t'} = s_\ell^{t'} = s_n^{t'} = \text{sign}(E^{t'})$ for all $t' \geq t$. Now one of three things happens.

- If $E^t + P_i^t > 0$, then $s_i^t = 1$. But this can only happen if $E^t > 0$, so that $s_n^t = s_j^t = s_k^t = s_\ell^t = 1$. Thus, we are in a unanimous positive deliberative equilibrium. But $E^t + P_i^t > 0$ if and only if $P_j^0 + P_k^0 + P_\ell^0 + P_i^0 > 0$, so this equilibrium is full-disclosure equivalent.
- If $P_i^t + E^t = 0$, then $s_i^t = 0$, and $E^t = |P_i^t|$. Again this implies $E^t > 0$, so that $s_n^t = s_j^t = s_k^t = s_\ell^t = 1$. In this case, i will disclose E^t further pieces of negative evidence, after which time all agents switch to neutral opinions, and we reach a neutral deliberative equilibrium. But $P_i^t + E^t = 0$ if and only if $P_j^0 + P_k^0 + P_\ell^0 + P_i^0 = 0$, so this equilibrium is full-disclosure equivalent.
- If $P_i^t + E^t < 0$, then $s_i^t = -1$, and $E^t < |P_i^t|$. If $E^t < 0$, then $s_j^t = s_k^t = s_\ell^t = s_n^t = -1$, so the group is immediately in a negative deliberative equilibrium. On the other hand, if $E^t \geq 0$, then $s_n^t = s_j^t = s_k^t = s_\ell^t \geq 0$. In this case, i will disclose $E^t + 1$ further pieces of negative evidence, after which time all agents switch to negative opinions, and we again reach a negative deliberative equilibrium. But $P_i^t + E^t < 0$ if and only if $P_j^0 + P_k^0 + P_\ell^0 + P_i^0 < 0$, so in either case, the negative equilibrium is full-disclosure equivalent.

Scenario 3. Suppose i runs out of private evidence before j, k and ℓ all run out of their evidence. Suppose this happens in round t. Then we must have $P_i^{t-1} = -1$ (i.e. i only had one remaining piece of evidence at time $t-1$), and $s_i^{t-1} < \text{Maj}^{t-1}$ (so that i dissented, and disclosed this evidence). There are two possibilities: either (a) $s_i^{t-1} = 0$ and $\text{Maj}^{t-1} = 1$, or (b) $s_i^{t-1} = -1$ and $\text{Maj}^{t-1} \geq 0$.

State	s_i^t	s_n^t	s_j^t	s_k^t	s_ℓ^t	Maj^t	Disclosers	E^{t+1} =?	Go to state...
A	-1	-1	-1	-1	-1	-1	$\emptyset$	E^t	equilibrium
B	-1	-1	-1	-1	0	-1	ℓ	E^t+1	B, D or F
C	-1	-1	-1	-1	1	-1	ℓ	E^t+1	C, E or G
D	-1	-1	-1	0	0	-1	k, ℓ	E^t+2	H, I, J, K, L or M
E	-1	-1	-1	0	1	-1	k, ℓ	E^t+2	H, I, J, K, L or M
F	-1	-1	0	0	0	-1	j, k, ℓ	E^t+3	J, K, L, M or N
G	-1	-1	0	0	1	-1	j, k, ℓ	E^t+3	J, K, L, M or N
H	-1	-1	-1	1	1	-1	k, ℓ	E^t+2	H, I, J, K, L or M
I	-1	-1	0	1	1	0	i; k, ℓ	E^t+1	J or K
J	-1	-1	1	1	1	1	i	E^t-1	F, G, I or J
K	-1	0	1	1	1	1	i	E^t-1	F, G, I or J
L	-1	1	1	1	1	1	i	E^t-1	K or L
M	0	1	1	1	1	1	i	E^t-1	M^*
N	1	1	1	1	1	1	$\emptyset$	E^t	equilibrium

Table 1: The proof of *Case (vii)*, Scenario 4.

(a) If $s_i^{t-1} = 0$ (and $\text{Maj}^{t-1} = 1$), then $E^{t-1} = 1$. After i discloses her last piece of evidence, we have $E^t = 0$ and $s_i^t = 0$; go to *Case (i)*.

(b) If $s_i^{t-1} = -1$ (and $\text{Maj}^{t-1} \geq 0$), then $E^{t-1} = 0$. Since (by hypothesis), at least one of j, k and ℓ still has private evidence, we must have $s_j^{t-1} = 1$, $s_k^{t-1} = 1$ or $s_\ell^{t-1} = 1$. In round t, after i has disclosed her last piece of negative evidence, we have $E^t = -1$, while $s_j^t \geq 0$, $s_k^t \geq 0$, or $s_\ell^t \geq 0$. Meanwhile, $s_i^t = s_n^t = -1$. There are now three subcases.

(b1) If $\text{Maj}^t = 1$, then we must have $s_j^t = s_k^t = s_\ell^t = 1$, and this is a deliberative equilibrium. Since $E^t = -1$, this can only occur if $P_j^t \geq 2$, $P_k^t \geq 2$ and $P_\ell^t \geq 2$, so this equilibrium is full-disclosure equivalent.

(b2) If $\text{Maj}^t = 0$, then we must have $s_j^t = 0$, while $s_k^t = s_\ell^t = 1$ (for some permutation of j, k and ℓ). Then k and ℓ dissent, so that $E^{t+1} = E^t + 2 = 1$ by statement (A1), so $s_i^{t+1} = s_n^{t+1} = s_j^{t+1} = s_k^{t+1} = s_\ell^{t+1} = 1$, which is a deliberative equilibrium. Since $E^t = -1$, subcase (b2) can only occur if $P_j^t \geq 1$, $P_k^t \geq 2$ and $P_\ell^t \geq 2$, so this equilibrium is full-disclosure equivalent.

(b3) Suppose $\text{Maj}^t = -1$. As already noted, $s_j^t \geq 0$, $s_k^t \geq 0$, or $s_\ell^t \geq 0$, so at least one of these agents will dissent and disclose positive evidence. If only one dissents, then we get $E^{t+1} = 0$ and $s_i^{t+1} = 0$, so we once again go to *Case (i)*. If two or more dissent, then $E^{t+1} \geq E^t + 2 = 1$ by statement (A1), so we reach a unanimously positive, full-disclosure equivalent deliberative equilibrium by the same logic as case (b2).

Scenario 4. Suppose that nobody ever runs out of private evidence during deliberation. In that case, as the agents disclose information, the group moves around between the states $\{A, B, C, \ldots, N\}$ shown in Table 1. Each "state" in this table describes a configuration of opinions, based on the assumption that $s_i^t \leq s_n^t \leq s_j^t \leq s_k^t \leq s_\ell^t$

(which is always true for a suitable permutation of j, k, and ℓ). Note that $s_n^t = 0$ if and only if $E^t = 0$, in which case we must have $s_i^t = -1$ and $s_j^t = s_k^t = s_\ell^t = 1$ (since we assume no one ever runs out of evidence). Thus, aside from state K, we have $s_n^t \neq 0$ in all states.

In each state, the majority opinion determines who will dissent. Since we assume that no agents ever run out of evidence, all dissenting agents always have evidence to disclose. (Except n, who never discloses anything, by Assumption (C).) Depending upon how this newly disclosed evidence interacts with the undisclosed private information of other agents, the group could transition into several different states from a given state. State transitions involving disclosure by only one agent must obey three principles:

(i) A single new piece of positive evidence will switch any zero opinion to positive, and it may or may not switch a negative opinion to zero. But it cannot switch a negative opinion to a positive opinion.

(ii) Likewise, a single new piece of negative evidence will switch any zero opinion to negative, and it may or may not switch a positive opinion to zero. But it cannot switch a positive opinion to a negative opinion.

(iii) If only a single agent discloses evidence, this may change other agents' opinions, but it cannot change her own opinion.

For example, in State B, agent ℓ discloses one piece of positive evidence. This could change the opinions of j or k from -1 to 0, but not to 1 (by principle (i)). In particular, it is impossible to move to state K. Since K is the only state in which $s_n^t = 0$, this means that ℓ cannot change n's opinion from -1 to 0. Finally, ℓ's disclosure cannot change ℓ's own opinion from 0 to 1 (by principle (iii)). Thus, the group will either remain in state B, or move to D or F.

However, if two or more agents disclose positive evidence simultaneously, then it is possible for them to switch even a negative opinion to a positive opinion, and if either of them had a neutral opinion, then it will switch to a positive opinion due to the *other* agent's newly disclosed evidence. For example, in State D, agents k and ℓ are both neutral, and *both* disclose positive evidence. Thus, they both switch to positive opinions, and also potentially move agents n and j to a neutral or positive opinion, and perhaps move i to a neutral opinion. (However, it is not possible that i moves to a positive opinion. This could only occur if $E^t = -1$ so that $E^{t+1} = 1$, while $P_i^t = 0$, violating the assumption that agents never run out of evidence.) Thus, the group can go to states H, I, J, K, L, or M.

Note that state M is special, since by principle (iii), the only possible transition from M is back into M itself. To understand why this does not lead to an infinite loop, note that the combination of $s_i^t = 0$ and $s_n^t = 1$ can only occur if $E^t > 0$ and $P_i^t = -E^t$. Thus, agent i will disclose one piece of negative evidence in each of the next E^t rounds until she runs out of evidence, which means we are back in Case (a) of *Scenario 3*. (Thus, since *Scenario 4* assumes no agent ever runs out of evidence, state M is actually impossible under this scenario. But we include it in Table 1

anyways, since this impossibility is not immediately obvious.)

Table 1 contains two possible equilibria: States A and N. But an inspection of Table 1 reveals that State A is unreachable from any other state. Thus, only N can occur as a terminal equilibrium of the deliberation. If N occurs at time t, then $E^t + P_i^t > 0$ (because $s_i^t = 1$). Meanwhile, P_j^t, P_k^t, and P_ℓ^t are all non-negative. Thus, $E^t + P_i^t + P_j^t + P_k^t + P_\ell^t > 0$, so the equilibrium is full-disclosure equivalent.

Case (viii). Suppose $|\mathcal{I}^+| = 3$, $|\mathcal{I}^-| = 1$, and $|\mathcal{N}| = 2$. Let $\mathcal{I}^- = \{i\}$, let $\mathcal{I}^+ = \{j, k, \ell\}$, and let $\mathcal{N} = \{n, m\}$. The argument is like *Case (iv)*. Deliberation unfolds as follows.

Round 0. $\text{Maj}^0 = -1$, so i dissents and discloses negative evidence, so $E^1 = -1$. (The agents in $\mathcal{N}$ do not disclose anything, by Assumption (C).)

Round 1. $E^1 = -1$. Thus, $s_i^1 = s_n^1 = s_m^1 = -1$, while $s_j^1 \geq 0$, $s_k^1 \geq 0$ and $s_\ell^1 \geq 0$. There are two cases.

- If one of s_j^1, s_k^1, or s_k^1 is zero, then $\text{Maj}^1 = -1$. In this case, j, k and ℓ all dissent and disclose positive evidence, so $E^2 = E^1 + 3 = 2$ by statement (A1); go to *Round 2* below.
- If $s_j^1 = s_k^1 = s_k^1 = 1$, then $\text{Maj}^1 = 0$. Once again, j, k and ℓ all dissent and disclose positive evidence. But now i also dissents. If $P_i^1 = 0$, then i has no evidence left to disclose, so $E^2 = E^1 + 3 = 2$ by statement (A1); go to *Round 2*. However, if $P_i^1 \leq -1$, then i discloses negative evidence, so that $E^2 = E^1 + 3 - 1 = 1$ by statement (A1). In this case, skip *Round 2* and go to *Round 3* below.

Round 2. $E^2 = 2$, so $s_n^2 = s_m^2 = s_j^2 = s_k^2 = s_\ell^2 = 1$, so $\text{Maj}^2 = 1$. There are two cases:

- If $P_i^2 \geq -1$, then $s_i^2 = 1$, so we reach a deliberative equilibrium that is full-disclosure equivalent.
- If $P_i^2 \leq -2$, then $s_i^2 = 0$ or -1, so i dissents and discloses negative evidence, so $E^3 = E^2 - 1 = 1$. Go to *Round 3*.

Round 3. $E^3 = 1$, so $s_n^3 = s_m^3 = s_j^3 = s_k^3 = s_\ell^3 = 1$, so $\text{Maj}^3 = 1$. There are two cases:

- If $P_i^3 = 0$, then $s_i^3 = 1$, so we reach a deliberative equilibrium that is full-disclosure equivalent.
- If $P_i^3 \leq -1$, then i dissents and discloses negative evidence, so $E^4 = E^3 - 1 = 0$. Go to *Round 4*.

Round 4. We now have $E^4 = 0$, just as in *Round 0*. There are six possible scenarios.

- If P_i^4, P_j^4, P_k^4 and P_ℓ^4 are all nonzero, then return to *Round 0* and repeat the argument found there.
- If $P_i^4 = 0$ while at least one of P_j^4, P_k^4 or P_ℓ^4 is nonzero, then we have entered *Case (i)* above.
- If $P_i^4 \leq -1$ while $P_j^4 = P_k^4 = P_\ell^4 = 0$, then we have entered *Case (ii)* above.
- If $P_k^4 = P_\ell^4 = 0$ while P_i^4 and P_j^4 are nonzero (for some permutation of j, k, ℓ), then we have entered *Case (iii)* above.

- If $P_\ell^4 = 0$ while P_i^4, P_j^4 and P_k^4 are nonzero (for some permutation of j, k, ℓ), then we have entered *Case (v)* above.
- If $P_i^4 = P_j^4 = P_k^4 = P_\ell^4 = 0$, then $\text{Maj}^4 = 0$, and this is a full-disclosure equivalent equilibrium.

The process described above eventually terminates when one or more agents runs out of evidence. In all terminal states, the deliberative equilibrium is full-disclosure equivalent.

Case (ix). Suppose $|\mathcal{I}^+| = 3$, $|\mathcal{I}^-| = 1$, and $|\mathcal{N}| \geq 3$. Let $\mathcal{I}^- = \{i\}$ and $\mathcal{I}^+ = \{j, k, \ell\}$. The argument is somewhat similar to *Case (v)*; we model the deliberation as a state-transition system. Here are the possible states.

Start. $E^0 = 0$, so that $s_n^0 = 0$ for all $n \in \mathcal{N}$. Meanwhile $P_i^0 \leq -1$, $P_j^0 \geq 1$, $P_k^0 \geq 1$ and $P_\ell^0 \geq 1$, so that $s_i^0 = -1$ while $s_j^0 = s_k^0 = s_\ell^0 = 1$; thus, $\text{Maj}^0 = 1$. Agent i dissents from the majority, and discloses a piece of negative evidence. (The agents in $\mathcal{N}$ do not disclose anything, by Assumption (C).) Thus, $E^1 = E^0 - 1 = -1$, so go to *State -1*.

State -1. $E^t = -1$, so that $s_n^t = -1$ for all $n \in \mathcal{N}$, and $s_i^t = -1$, so that $\text{Maj}^t = -1$. Meanwhile, $P_j^t \geq 1$, $P_k^t \geq 1$, and $P_\ell^0 \geq 1$, so that $s_j^t \geq 0$, $s_k^t \geq 0$, and $s_\ell^t \geq 0$. So j, k and ℓ dissent from the majority and disclose positive evidence. Thus, $E^{t+1} = E^t + 3 = 2$ by statement (A1), so go to *State 2*.

State 2. $E^t = 2$, so $s_j^t = s_k^t = s_\ell^t = 1$ and $s_n^t = 1$ for all $n \in \mathcal{N}$; thus $\text{Maj}^t = 1$. There are two subcases: either $P_i^t \geq -1$, or $P_i^t \leq -2$.

- If $P_i^t \geq -1$, then $s_i^t = 1$ (because $E^t + P_i^t \geq 1$). Thus, the group reaches unanimous deliberative equilibrium. Since $E^t + P_i^t \geq 1$, while $P_j^t \geq 0$, $P_k^t \geq 0$ and $P_\ell^t \geq 0$, this equilibrium is full-disclosure equivalent.
- If $P_i^t \leq -2$, then $s_i^t = 0$ or -1, so i dissents from the majority and discloses negative evidence. Thus, $E^{t+1} = E^t - 1 = 1$, so go to *State 1*.

State 1. $E^t = 1$, so $s_j^t = s_k^t = s_\ell^t = 1$ and $s_n^t = 1$ for all $n \in \mathcal{N}$; thus $\text{Maj}^t = 1$. There are two subcases: either $P_i^t = 0$, or $P_i^t \leq -1$.

- If $P_i^t = 0$, then $s_i^t = 1$. Thus, the group reaches unanimous deliberative equilibrium. Since $E^t + P_i^t = 1$, while $P_j^t \geq 0$, $P_k^t \geq 0$ and $P_\ell^t \geq 0$, this equilibrium is full-disclosure equivalent.
- If $P_i^t \leq -1$, then $s_i^t = 0$ or -1, so i dissents from the majority and discloses negative evidence. Thus, $E^{t+1} = E^t - 1 = 0$. Go to *State 0*.

State 0. $E^t = 0$, so $s_j^t, s_k^t, s_\ell^t \geq 0$, while $s_n^t = 0$ for all $n \in \mathcal{N}$. There are five sub-cases.

- If $P_i^t = 0$, then go to *Case (i)*.
- If $P_i^t \leq -1$ while $P_j^t = P_k^t = P_\ell^t = 0$, then go to *Case (ii)*.
- If $P_i^t \leq -1$ and $P_j^t \geq 1$, while $P_k^t = P_\ell^t = 0$ (for some permutation of j, k, ℓ), then go to *Case (iii)*.

- If $P_i^t \leq -1$, $P_j^t \geq 1$ and $P_k^t \geq 1$, while $P_\ell^t = 0$ (for some permutation of j, k, ℓ), then go to *Case (v).*
- If $P_i^t \leq -1$, $P_j^t \geq 1$, $P_k^t \geq 1$, and $P_\ell^t \geq 1$, then go back to *Start.*

In all equilibria of this state-transition system, the decision is full-disclosure equivalent.

Case (x). If $|\mathcal{I}^-| = 2$ and $|\mathcal{I}^+| = 1$, then the analysis is similar to *Cases (iv)* and *(v)*, depending on whether $|\mathcal{N}| = 1$ or $|\mathcal{N}| \geq 2$.

Case (xi). If $|\mathcal{I}^-| = 3$ and $|\mathcal{I}^+| = 1$, then the analysis is similar to *Cases (vii), (viii)* and *(ix)*, depending on whether $|\mathcal{N}| = 1$, $|\mathcal{N}| = 2$, or $|\mathcal{N}| \geq 3$. □

Proof of Theorem 3. *Case 1.* Suppose that $E^0 = 0$. There are four subcases.

Case 1(a) Suppose $\mathcal{I}_0^0 = \mathcal{I}$. Then all agents already agree with the decision $D^0 = 0$ from formula (3D), so deliberative equilibrium is reached immediately. But

$$\sum_{m\in\mathcal{M}} y_m \underset{(*)}{=} \sum_{m\in\mathcal{M}\setminus\mathcal{C}^0} y_m \underset{(\dagger)}{=} \sum_{i\in\mathcal{I}_0^0}\sum_{m\in\mathcal{M}_i} y_m = \sum_{i\in\mathcal{I}_0^0} 0 = 0,$$

where $(*)$ is because $E^0 = 0$, and $(\dagger)$ is by substituting $\mathcal{I} = \mathcal{I}_0^0$ into Lemma A1(a). Thus, this neutral decision is the same decision that would have been reached with full disclosure; hence the equilibrium is full-disclosure equivalent.

Case 1(b) Suppose $\mathcal{I}_-^0 = \emptyset$ while $\mathcal{I}_+^0 \neq \emptyset$. Then someone in $\mathcal{I}_+^0$ will disclose a piece of positive evidence during the first round of deliberation. At this point, all agents in $\mathcal{I}$ will agree with the decision $D^1 = 1$ from formula (3D). Thus, deliberative equilibrium is reached in Round 1. But

$$\sum_{m\in\mathcal{M}} y_m \underset{(*)}{=} \sum_{m\in\mathcal{M}\setminus\mathcal{C}^0} y_m \underset{(\dagger)}{=} \sum_{i\in\mathcal{I}_+^0}\sum_{m\in\mathcal{M}_i} y_m > 0,$$

where $(*)$ is because $E^0 = 0$, and $(\dagger)$ is by substituting $\mathcal{I} = \mathcal{I}_+^0 \cup \mathcal{I}_0^0$ into Lemma A1(a). Thus, this positive decision is the same decision that would have been reached with full disclosure; hence the equilibrium is full-disclosure equivalent.

Case 1(c) Likewise, if $\mathcal{I}_+^0 = \emptyset$ while $\mathcal{I}_-^0 \neq \emptyset$, then the group will converge to equilibrium with $D^1 = -1$ after one round of deliberation, and this is full-disclosure equivalent.

Case 1(d) Finally, suppose $\mathcal{I}_+^0 \neq \emptyset$ and $\mathcal{I}_-^0 \neq \emptyset$. During the first round of deliberation, either someone in $\mathcal{I}_+^0$ will disclose a piece of positive evidence, or someone in $\mathcal{I}^-$ will disclose a piece of negative evidence. At this point, $E^1 \neq 0$, so starting in round 2, we can invoke the argument from *Case 2.* below.

Case 2. Suppose that $E^0 \neq 0$. Without loss of generality, suppose $E^0 < 0$ (the other case is similar). By hypothesis, there is initial disagreement; thus, $\mathcal{I}_+^0 \cup \mathcal{I}_0^0 \neq \emptyset$. But if i in $\mathcal{I}_+^0 \cup \mathcal{I}_0^0$, then

$$P_i^0 + E^0 \quad \geq \quad 0. \tag{A3}$$

Agent i dissents from the (negative) opinion derived from the publicly available evidence $\mathcal{C}^0$. So during the first round, she will disclose a piece of (positive) evidence. As a result, $E^1 = E^0 + 1$, while $P_i^1 = P_i^0 - 1$. Thus, from inequality (A3) we obtain

$$P_i^1 + E^1 \quad \geq \quad 0. \tag{A4}$$

Thus, i's opinion remains non-negative. Furthermore, inequality (A4) clearly holds for all of the other members of $\mathcal{I}_+^0 \cup \mathcal{I}_0^0$; thus we have $\mathcal{I}_+^0 \cup \mathcal{I}_0^0 \subseteq \mathcal{I}_+^1 \cup \mathcal{I}_0^1$.

Now, if $E^1 < 0$, then during the second round of deliberation, some member j of $\mathcal{I}_+^1 \cup \mathcal{I}_0^1$ will disclose another piece of positive evidence, so that $E^2 = E^1 + 1$, while $P_j^2 = P_j^1 - 1$. Thus,

$$P_k^2 + E^2 \quad \geq \quad 0 \quad \text{for all } k \in \mathcal{I}_+^1 \cup \mathcal{I}_0^1, \text{ and hence } \quad \mathcal{I}_+^1 \cup \mathcal{I}_0^1 \quad \subseteq \quad \mathcal{I}_+^2 \cup \mathcal{I}_0^2.$$

Let $T_0 := |E^0|$. By inductively repeating the above argument for each of rounds $t = 1, 2, \ldots, T_0$, we see that during each round, one of the dissenting agents in $\mathcal{I}_+^t \cup \mathcal{I}_0^t$ will discloses a piece of positive evidence, after which time

$$P_i^{t+1} + E^{t+1} \quad \geq \quad 0 \quad \text{for all } i \in \mathcal{I}_+^t \cup \mathcal{I}_0^t, \text{ and hence } \quad \mathcal{I}_+^t \cup \mathcal{I}_0^t \quad \subseteq \quad \mathcal{I}_+^{t+1} \cup \mathcal{I}_0^{t+1}.$$

This process will continue until round T_0, at which point $E^{T_0} = 0$.

Now, let $t \geq T_0$. There are three possible disequilibrium scenarios at time t.

- If $E^t = 0$, then go back to *Case 1.*
- Suppose $E^t = 1$. If $\mathcal{I}_0^t \cup \mathcal{I}_-^t \neq \emptyset$, then someone in $\mathcal{I}_0^t \cup \mathcal{I}_-^t$ will disclose negative evidence, so that $E^{t+1} = 0$.
- Suppose $E^t = -1$. If $\mathcal{I}_0^t \cup \mathcal{I}_+^t \neq \emptyset$, then someone in $\mathcal{I}_0^t \cup \mathcal{I}_+^t$ will disclose positive evidence, so that $E^{t+1} = 0$.

Note that for all $t \geq T_0$, we have $E^t \in \{-1, 0, 1\}$. This process will continue until a time $T \geq T_0$ when deliberative equilibrium is reached. There are then three subcases.

Case 2(a). If $E^T = 0$, then we must have $\mathcal{I}_+^T \cup \mathcal{I}_-^T = \emptyset$. In other words $\mathcal{I} = \mathcal{I}_0^T$, so the consensus opinion is neutral. Equation (3B) yields $P_i^T = 0$ for all $i \in \mathcal{I}$; thus

$$E^T + \sum_{i \in \mathcal{I}} P_i^T \quad = \quad 0, \tag{A5}$$

because each of the summands is zero. Combining equation (A5) with Lemma A1(b), we conclude that $\sum_{m \in \mathcal{M}} y_m = 0$. Thus, this equilibrium is full-disclosure equivalent.

Case 2(b). If $E^T = 1$, then we must have $\mathcal{I}_0^T \cup \mathcal{I}_-^T = \emptyset$. In other words $\mathcal{I} = \mathcal{I}_+^T$, so the consensus opinion is positive. Equation (3B) yields $P_i^T \geq 0$ for all $i \in \mathcal{I}$. Thus,

$$E^T + \sum_{i \in \mathcal{I}} P_i^T \quad > \quad 0, \tag{A6}$$

because each of the summands is non-negative, and $E^T > 0$. Combining equation (A6) with Lemma A1(b) yields $\sum_{m \in \mathcal{M}} y_m > 0$. So this equilibrium is full-disclosure equivalent.

Case 2(c). Likewise, if $E^T = -1$, then $\mathcal{I}_0^T \cup \mathcal{I}_+^T = \emptyset$; hence $\mathcal{I} = \mathcal{I}_-^T$, so that the consensus opinion is negative. Equation (3B) yields $P_i^T \leq 0$ for all $i \in \mathcal{I}$. Thus,

$$E^T + \sum_{i \in \mathcal{I}} P_i^T \quad < \quad 0, \tag{A7}$$

because each of the summands is non-positive, and $E^T < 0$. Combining equation (A7) with Lemma A1(b) yields $\sum_{m \in \mathcal{M}} y_m < 0$. So this equilibrium is full-disclosure equivalent. □

The proof of Theorem 4 depends on three lemmas. For any $t \in \mathbb{N}$, let $\mathcal{N}^t := \{i \in \mathcal{I}; P_i^t = 0\}$. (Thus, $\mathcal{N}^0$ is the set of internally neutral agents.)

Lemma A2 *For any $t \in \mathbb{N}$ and $i \in \mathcal{N}^t$, i never discloses any further evidence after time t, in either the serial or parallel advisory protocols. Thus, $\mathcal{P}_i^{t'} = \mathcal{P}_i^t$ for all $t' \geq t$. It follows that $\mathcal{N}^0 \subseteq \mathcal{N}^1 \subseteq \mathcal{N}^2 \subseteq \cdots$.*

Proof: For all $i \in \mathcal{N}^t$, we have $s_i^t \underset{(*)}{=} \text{sign}(E^t) \underset{(\dagger)}{=} D^t$, where $(*)$ is by formula (3B) and $(\dagger)$ is by formula (3D). Thus i has no incentive to disclose any further information. Thus, $\mathcal{P}_i^{t+1} = \mathcal{P}_i^t$, so that $i \in \mathcal{N}^{t+1}$. Inductively, $\mathcal{P}_i^{t'} = \mathcal{P}_i^t$, and hence $i \in \mathcal{N}^{t'}$ for all $t' \geq t$. □

Lemma A3 *Suppose there exists some $t \in \mathbb{N}$ such that $E^t = 0$, and either $|\mathcal{I}_+^t| = 1$ or $|\mathcal{I}_-^t| = 1$. Then in the parallel advisory protocol, the deliberative equilibrium is full-disclosure equivalent.*

Proof: Since $E^t = 0$, we have $\mathcal{I}_0^t = \mathcal{N}^t$. Thus, Lemma A2 says that agents in $\mathcal{I}_0^t$ will never disclose any further evidence after time t.

Suppose that $|\mathcal{I}_+^t| = 1$. (The argument when $|\mathcal{I}_-^t| = 1$ is similar.) Let $\mathcal{I}_+^t = \{j\}$. Suppose we reach deliberative equilibrium in some round $T \geq t$. Observe that

$$E^T \quad = \quad E^t + P - N \quad = \quad P - N, \tag{A8}$$

where P is the total positive evidence disclosed by j after round t, and N is the total negative evidence disclosed by everyone in $\mathcal{I}_-^t$ after round t. (As already noted, agents in $\mathcal{I}_0^t$ will never disclose any evidence after round t.) There are now three cases.

Case 1. If $D^T = 1$, then all agents in $\mathcal{I}^t_-$ must have exhausted their negative evidence. Thus, we must have $N = \sum_{i \in \mathcal{I}^t_-} |P^t_i|$, while $P \leq P^t_j$. Thus,

$$\begin{aligned}\sum_{m \in \mathcal{M}} y_m &\underset{(*)}{=} E^t + \sum_{i \in \mathcal{I}} P^t_i \quad = \quad 0 + P^t_j + \sum_{i \in \mathcal{I}^t_-} P^t_i + \sum_{i \in \mathcal{I}^t_0} P^t_i \\ &= P^t_j + \sum_{i \in \mathcal{I}^t_-} P^t_i + \sum_{i \in \mathcal{I}^t_0} 0 \quad \geq \quad P - N \underset{(\dagger)}{=} E^T \underset{(\diamond)}{>} 0,\end{aligned}$$

where $(*)$ is by Lemma A1(b), $(\dagger)$ is by equation (A8), and $(\diamond)$ is by equation (3D), because $D^T = 1$. Thus, the decision with full disclosure would also have been positive, so the equilibrium is full-disclosure equivalent.

Case 2. If $D^T = -1$, then j must have exhausted her positive evidence. Thus, we must have $P = P^t_j$, while $N \leq \sum_{i \in \mathcal{I}^t_-} |P^t_i|$. Equivalently, $-N \geq \sum_{i \in \mathcal{I}^t_-} P^t_i$. Thus,

$$\begin{aligned}\sum_{m \in \mathcal{M}} y_m &\underset{(*)}{=} E^t + \sum_{i \in \mathcal{I}} P^t_i \quad = \quad 0 + P^t_j + \sum_{i \in \mathcal{I}^t_-} P^t_i + \sum_{i \in \mathcal{I}^t_0} P^t_i \\ &= P^t_j + \sum_{i \in \mathcal{I}^t_-} P^t_i + \sum_{i \in \mathcal{I}^t_0} 0 \quad \leq \quad P - N \underset{(\dagger)}{=} E^T \underset{(\diamond)}{<} 0,\end{aligned}$$

where $(*)$ is by Lemma A1(b), $(\dagger)$ is by equation (A8), and $(\diamond)$ is by equation (3D), because $D^T = -1$. Thus, the decision with full disclosure would also have been negative, so the equilibrium is full-disclosure equivalent.

Case 3. If $D^T = 0$, then both sides must have exhausted their evidence. Thus, we must have $P = P^t_j$ and $N = \sum_{i \in \mathcal{I}^t_-} |P^t_i|$. Thus,

$$\begin{aligned}\sum_{m \in \mathcal{M}} y_m &\underset{(*)}{=} E^t + \sum_{i \in \mathcal{I}} P^t_i \quad = \quad 0 + P^t_j + \sum_{i \in \mathcal{I}^t_-} P^t_i + \sum_{i \in \mathcal{I}^t_0} P^t_i \\ &= P^t_j + \sum_{i \in \mathcal{I}^t_-} P^t_i + \sum_{i \in \mathcal{I}^t_0} 0 \quad = \quad P - N \underset{(\dagger)}{=} E^T \underset{(\diamond)}{=} 0,\end{aligned}$$

where $(*)$ is by Lemma A1(b), $(\dagger)$ is by equation (A8), and $(\diamond)$ is by equation (3D), because $D^T = 0$. Thus, the decision with full disclosure would also have been neutral, so the equilibrium is full-disclosure equivalent. □

Lemma A4 *Suppose there is some time $t \in \mathbb{N}$ such that $E^t = 0$, and such that either $\mathcal{I}^t_+ = \emptyset$ or $\mathcal{I}^t_- = \emptyset$. Then in the parallel advisory protocol, the deliberative equilibrium is full-disclosure equivalent.*

Proof: Suppose $\mathcal{I}_-^t = \emptyset$. (The argument when $\mathcal{I}_+^t = \emptyset$ is similar.) If also $\mathcal{I}_+^t = \emptyset$, then we must have $\mathcal{I} = \mathcal{I}_0^t$; thus, the group has already reached deliberative equilibrium at time t, with $D^t = 0$. However,

$$\sum_{m \in \mathcal{M}} y_m \underset{(*)}{=} E^t + \sum_{i \in \mathcal{I}} P_i^t \underset{(\dagger)}{=} 0 + \sum_{i \in \mathcal{I}} 0 = 0,$$

where $(*)$ is by Lemma A1(b), and $(\dagger)$ is because $\mathcal{I} = \mathcal{I}_0^t$ and $E^t = 0$ by hypothesis. Thus, the decision with full disclosure would also be neutral. Thus, the deliberative equilibrium is full-disclosure equivalent.

Now suppose $\mathcal{I}_+^t \neq \emptyset$. Then

$$\sum_{m \in \mathcal{M}} y_m \underset{(*)}{=} E^t + \sum_{i \in \mathcal{I}} P_i^t = 0 + \sum_{i \in \mathcal{I}_0^t} P_i^t + \sum_{i \in \mathcal{I}_+^t} P_i^t \underset{(\dagger)}{>} 0,$$

where $(*)$ is by Lemma A1(b), and $(\dagger)$ is because each of the terms in the first sum is zero, while each term in the second sum is positive. Thus, the decision with full disclosure would be positive.

During round $t+1$ of deliberation, each agent in $\mathcal{I}_+^t$ discloses one piece of positive information. Since $E^t = 0$, we have $\mathcal{I}_0^t = \mathcal{N}^t$. Thus, Lemma A2 says that agents in $\mathcal{I}_0^t$ do not discloses any evidence at time $t+1$. Finally, no one discloses any negative information, because $\mathcal{I}_-^t = \emptyset$. Thus, we have $E^{t+1} > 0$, so $D^{t+1} = 1$ by equation (3D). At this point, for every $i \in \mathcal{I}_+^t$, we have $E^{t+1} + P_i^{t+1} > P_i^{t+1} \geq 0$, so equation (3B) implies that all the agents who were in $\mathcal{I}_+^t$ retain a positive opinion at time $t+1$.

Meanwhile, for every $i \in \mathcal{I}_0^t$, we have $E^{t+1} + P_i^{t+1} = E^{t+1} > 0$, so equation (3B) implies that all the agents who were in $\mathcal{I}_0^t$ now also have a positive opinion at time $t+1$. By hypothesis, $\mathcal{I}_-^t = \emptyset$, so these two cases account for everyone in $\mathcal{I}$. Thus, everyone in $\mathcal{I}$ has a positive opinion at time $t+1$, so deliberative equilibrium is obtained with $D^{t+1} = 1$ —a deliberative equilibrium which agrees with the full-disclosure decision. □

Proof of Theorem 4. If $|\mathcal{I}_+^0| = 1$ or $|\mathcal{I}_-^0| = 1$, then Lemma A3 says that the deliberative equilibrium is full-disclosure equivalent. So, we can assume without loss of generality that $|\mathcal{I}_+^0| \geq 2$ and $|\mathcal{I}_-^0| \geq 2$. Since $|\mathcal{I}_+^0| + |\mathcal{I}_-^0| \leq |\mathcal{I}^*| \leq 4$, it follows that $|\mathcal{I}_+^0| = |\mathcal{I}_-^0| = 2$ and hence $|\mathcal{I}^*| = 4$. Let $\mathcal{I}_+^0 = \{j, k\}$ and let $\mathcal{I}_-^0 = \{\ell, m\}$. For all $i \in \mathcal{I}^*$, let $P_i^0 := \sum_{m \in \mathcal{M}_i} y_m$; thus, $P_j^0, P_k^0 > 0 > P_\ell^0, P_m^0$. Let $T_0 := \min\{P_j^0, P_k^0, |P_\ell^0|, |P_m^0|\}$.

Since $D^0 = 0$, all four agents will disclose evidence during the first round of deliberation. But since $|\mathcal{I}_+^0| = |\mathcal{I}_-^0|$, the result will be that $D^1 = 0$, by statement (A1) and equation (3D). Thus, during the second round, all four agents will *again* disclose evidence, leading again to $D^2 = 0$ by (A1) and (3D). Inductively, all four agents will disclose evidence in *each* of the rounds $t = 1, 2, \ldots, T_0$. In round t, we will have

$$E^t = 0 \text{ by (A1)}, \quad P_i^t = P_i^0 - t \text{ for both } i \in \mathcal{I}_+^0, \text{ and } P_i^t = P_i^0 + t \text{ for both } i \in \mathcal{I}_-^0.$$

Once we get to round T_0, we will thus have $P_i^{T_0} = 0$ for some $i \in \mathcal{I}^*$

At this point, there are at most three agents remaining with non-neutral opinions. There are now four cases.

Case 1. If no non-neutral agents remain, or only one non-neutral agent remains, then either $\mathcal{I}_+^0 = \emptyset$ or $\mathcal{I}_-^0 = \emptyset$ (or both), so Lemma A4 says the equilibrium is full-disclosure equivalent.

Case 2. If only two non-neutral agents remain, and they both belong to $\mathcal{I}_+^0$ or they both belong to $\mathcal{I}_-^0$, then Lemma A4 says the equilibrium is full-disclosure equivalent.

Case 3. If only two non-neutral agents remain, and one belongs to $\mathcal{I}_+^0$ while the other belongs to $\mathcal{I}_-^0$, then Lemma A3 says the equilibrium is full-disclosure equivalent.

Case 4. If three non-neutral agents remain, then we have either two negative agents and one positive agent, or two positive agents and one negative agent. Either way, Lemma A3 says that the deliberative equilibrium is full-disclosure equivalent. □

B Proofs from Section 4

Proof of Proposition 4.1. Suppose $\mathcal{C}_i^t = \mathcal{M}_i$ for all $i \in \mathcal{I}$ —in other words, each agent has revealed *all* of her evidence at time t. Then $\mathcal{C}^t = \mathcal{M}$ (by simply comparing equations (2C) and (2D)). Let π denote the common prior. For all $i \in \mathcal{I}$, we have $\mathcal{P}_i^t = \emptyset$, so formula (2E) yields

$$B_i^t(x) \quad = \quad \underset{\text{(SNC)}}{\frac{\pi(x)}{}} \cdot \prod_{c \in \mathcal{C}^t} \rho_x(y_c), \qquad \text{for all } x \in \mathcal{X}.$$

Thus, all agents have the same beliefs —hence they are in deliberative equilibrium. □

Theorem 5 is just the special case of Theorem 6 when $\pi_i = \nu$ for all $i \in \mathcal{I}$. So it suffices to prove Theorem 6.

Proof of Theorems 5 and 6. Rather than working with a conditional probability distribution like $P_{\mathbf{y}}$ in equation (2A), it will be more convenient for us to work with the corresponding *log likelihood ratio matrix*. This is the antisymmetric $\mathcal{X} \times \mathcal{X}$ matrix $\mathbf{B}^{\mathbf{y}} = [b_{w,x}^{\mathbf{y}}]_{w,x \in \mathcal{X}}$, where, for all $w, x \in \mathcal{X}$,

$$b_{w,x}^{\mathbf{y}} \;:=\; \log\left(\frac{P_{\mathbf{y}}(w)}{P_{\mathbf{y}}(x)}\right) \underset{(*)}{=} \log[\pi(w)] + \sum_{m=1}^{M} \log[\rho_w(y_m)] - \log[\pi(x)] - \sum_{m=1}^{M} \log[\rho_x(y_m)], \quad \text{(B1)}$$

where $(*)$ is obtained by substituting in the expression (2A). (Note that these logarithms are always finite, because we have assumed that $\pi \in \Delta^*(\mathcal{X})$ and $\rho_x \in \Delta^*(\mathcal{Y})$ for all

$x \in \mathcal{X}$.) One advantage of log likelihood ratios is that they eliminate the need for the normalization constants we have been denoting by "(SNC)". Another advantage is that they turn multiplicative expressions like (2A), (2B), (2E), (4B), (4D) and (4E) into more transparent additive expressions.

Let $\mathbf{B}^\emptyset = [b^\emptyset_{w,x}]_{w,x\in\mathcal{X}}$ be the antisymmetric $\mathcal{X} \times \mathcal{X}$ matrix defined by setting $b^\emptyset_{w,x} := \log[\pi(w)] - \log[\pi(x)]$, for all $w, x \in \mathcal{X}$. Meanwhile, for any $y \in \mathcal{Y}$, define the antisymmetric $\mathcal{X} \times \mathcal{X}$ matrix $\mathbf{E}^y = [e^y_{w,x}]_{w,x\in\mathcal{X}}$ by setting $e^y_{w,x} := \log[\rho_w(y)] - \log[\rho_x(y)]$, for all $w, x \in \mathcal{X}$. Then equation (B1) can be rewritten:

$$\mathbf{B}^{\mathbf{y}} \;=\; \mathbf{B}^\emptyset + \sum_{m=1}^{M} \mathbf{E}^{y_m}. \tag{B2}$$

In other words, the log likelihood ratio matrix $\mathbf{B}^{\mathbf{y}}$ representing posterior beliefs is the *sum* of a matrix $\mathbf{B}^\emptyset$ representing the prior beliefs and the matrices $\mathbf{E}^{y_1}, \ldots, \mathbf{E}^{y_M}$ representing the evidence received. This representation will be convenient to us.

For all $i \in \mathcal{I}$, let $\mathbf{B}^0_i$ be a log likelihood ratio matrix representing i's beliefs at time zero, before deliberation begins. In other words, this is the log likelihood ratio matrix of the probability distribution B^0_i in equation (2B). But we can also compute it directly, as follows. Let $\mathbf{B}^\emptyset_i$ be a log likelihood ratio matrix representing i's prior π_i. Then $\mathbf{B}^0_i$ is obtained by combining $\mathbf{B}^\emptyset_i$ with her initial information, via equation (B2):

$$\mathbf{B}^0_i \quad = \quad \mathbf{B}^\emptyset_i + \sum_{m\in\mathcal{M}_i} \mathbf{E}^{y_m}. \tag{B3}$$

Let $\mathbf{B}^t_i$ be the log likelihood ratio matrix representing agent i's beliefs B^t_i at time t, as defined by equation (2E). This is a combination of her prior, the publicly available evidence, and her own (undisclosed) private evidence. By applying (B2), we obtain:

$$\begin{aligned}
\mathbf{B}^t_i \;&=\; \mathbf{B}^\emptyset_i + \sum_{p\in\mathcal{P}^t_i} \mathbf{E}^{y_p} + \sum_{c\in\mathcal{C}^t} \mathbf{E}^{y_c} && \text{(B4)}\\
&=\; \mathbf{B}^\emptyset_i + \sum_{p\in\mathcal{P}^t_i} \mathbf{E}^{y_p} + \sum_{c\in\mathcal{C}^t_i} \mathbf{E}^{y_c} + \sum_{m\in\mathcal{C}^t\setminus\mathcal{C}^t_i} \mathbf{E}^{y_m}\\
&\underset{(*)}{=}\; \mathbf{B}^\emptyset_i + \sum_{m\in\mathcal{M}_i} \mathbf{E}^{y_m} + \sum_{c\in\mathcal{C}^t\setminus\mathcal{C}^t_i} \mathbf{E}^{y_c} && \text{(B5)}
\end{aligned}$$

where $(*)$ is because $\mathcal{M}_i = \mathcal{P}^t_i \sqcup \mathcal{C}^t_i$.

Assumption (B′) says there is an internally neutral agent n. Let $\mathbf{N}$ be the log likelihood ratio matrix of n's prior ν. Let B^t_n denote her beliefs at time t, as in equation (4D), and let $\mathbf{B}^t_n$ be the log likelihood ratio matrix of B^t_n. By applying (B2), we get

$$\mathbf{B}^t_n \quad = \quad \mathbf{N} + \sum_{c\in\mathcal{C}^t} \mathbf{E}^{y_c}. \tag{B6}$$

In deliberative equilibrium, all agents must agree with all other agents, and hence with the internally neutral n. Conversely, if all non-neutral agents agree with n, then they automatically agree with each other. So in groups with an internally neutral agent n, deliberative equilibrium is equivalent to requiring that $\mathbf{B}_i^t = \mathbf{B}_n^t$ for all $i \in \mathcal{I}$. For any $i \in \mathcal{I}$, we can rewrite (B6) as

$$\mathbf{B}_n^t \quad = \quad \mathbf{N} + \sum_{c \in \mathcal{C}_i^t} \mathbf{E}^{y_c} + \sum_{c \in \mathcal{C}^t \setminus \mathcal{C}_i^t} \mathbf{E}^{y_c}. \tag{B7}$$

We deduce that

$$\begin{aligned} \left(\mathbf{B}_n^t = \mathbf{B}_i^t\right) \quad &\underset{(*)}{\iff} \quad \left(\mathbf{N} + \sum_{c \in \mathcal{C}_i^t} \mathbf{E}^{y_c} + \sum_{c \in \mathcal{C}^t \setminus \mathcal{C}_i^t} \mathbf{E}^{y_c} = \mathbf{B}_i^{\emptyset} + \sum_{m \in \mathcal{M}_i} \mathbf{E}^{y_m} + \sum_{c \in \mathcal{C}^t \setminus \mathcal{C}_i^t} \mathbf{E}^{y_c}\right) \\ &\iff \quad \left(\sum_{c \in \mathcal{C}_i^t} \mathbf{E}^{y_c} = (\mathbf{B}_i^{\emptyset} - \mathbf{N}) + \sum_{m \in \mathcal{M}_i} \mathbf{E}^{y_m}\right), \end{aligned} \tag{B8}$$

where $(*)$ is obtained by substituting (B5) and (B7). There are now two cases: either $\{\mathcal{M}_i\}_{i \in \mathcal{I}}$ are mutually disjoint, or they overlap.

Case **(a)** Suppose that $\{\mathcal{M}_i\}_{i \in \mathcal{I}}$ are all disjoint from one another. (So Assumption (A) is trivially satisfied.) Recall that (by definition), they are also disjoint from the initial public information set $\mathcal{C}^0$. Then

$$\begin{aligned} \mathbf{B}_n^t \quad &\underset{(*)}{=} \quad \mathbf{N} + \sum_{c \in \mathcal{C}^t} \mathbf{E}^{y_c} \underset{(\dagger)}{=} \mathbf{N} + \sum_{c \in \mathcal{C}^0} \mathbf{E}^{y_c} + \sum_{i \in \mathcal{I}} \sum_{c \in \mathcal{C}_i^t} \mathbf{E}^{y_c} \\ &\underset{(\diamond)}{=} \quad \mathbf{N} + \sum_{c \in \mathcal{C}^0} \mathbf{E}^{y_c} + \sum_{i \in \mathcal{I}} (\mathbf{B}_i^{\emptyset} - \mathbf{N}) + \sum_{i \in \mathcal{I}} \sum_{m \in \mathcal{M}_i} \mathbf{E}^{y_m} \\ &\underset{(\ddagger)}{=} \quad \mathbf{N} + \sum_{i \in \mathcal{I}} (\mathbf{B}_i^{\emptyset} - \mathbf{N}) + \sum_{m \in \mathcal{M}} \mathbf{E}^{y_m} \end{aligned} \tag{B9}$$

Here, $(*)$ is by (B6), while $(\dagger)$ is by defining formula (2D), and the fact that the sets $\{\mathcal{C}_i^t\}_{i \in \mathcal{I}}$ must all be disjoint from one another and from $\mathcal{C}^0$. Next, $(\diamond)$ is by applying the equation in statement (B8) for all $i \in \mathcal{I}$. Finally $(\ddagger)$ is by equation (2C).

Let $\mathbf{M}$ be the log likelihood ratio matrix of the belief μ obtained from applying the multiplicative pooling rule to the beliefs $\{\pi_i\}_{i \in \mathcal{I}}$ with reference measure ν, as in equation (4E). Then $\mathbf{M} = \mathbf{N} + \sum_{i \in \mathcal{I}} (\mathbf{B}_i^{\emptyset} - \mathbf{N})$. Meanwhile, let $\mathbf{A} = [a_{w,x}]_{w,x \in \mathcal{X}}$ be the antisymmetric matrix defined by $a_{w,x} = \log[\alpha(w)/\alpha(x)]$, where α is defined as in (4A). Then

$$\mathbf{A} \quad = \quad \sum_{m \in \mathcal{M}} \mathbf{E}^{y_m}. \tag{B10}$$

Thus, equation (B9) says that $\mathbf{B}_n^t = \mathbf{M} + \mathbf{A}$. In other words, the probabilistic beliefs of the internally neutral agent have the form $B_n^t = \mu \cdot \alpha$, as claimed.

Case **(b)** Suppose that $\{\mathcal{M}_i\}_{i\in\mathcal{I}}$ are *not* disjoint from one another. For all $i \in \mathcal{I}$, define

$$\widetilde{\mathcal{M}}_i \;:=\; \mathcal{M}_i \setminus \left(\bigcup_{\substack{j\in\mathcal{I}\\ j\neq i}} \mathcal{M}_j\right) \quad \text{and} \quad \widehat{\mathcal{M}}_i \;:=\; \mathcal{M}_i \cap \left(\bigcup_{\substack{j\in\mathcal{I}\\ j\neq i}} \mathcal{M}_j\right) \;=\; \mathcal{M}_i \setminus \widetilde{\mathcal{M}}_i.$$

Thus, $\widetilde{\mathcal{M}}_i$ is the part of i's evidence which is *not* shared by any other agent, while $\widehat{\mathcal{M}}_i$ is the part of i's evidence that she shares with at least one other agent. Now define

$$\widetilde{\mathcal{C}}^0 \;:=\; \mathcal{C}^0 \cup \bigcup_{m\in\mathcal{M}} \widehat{\mathcal{M}}_m.$$

Any deliberative equilibrium satisfying Assumption (A) can be obtained through the following two-step procedure:

1. First, every agent discloses all evidence in $\widehat{\mathcal{M}}_i$ (so that Assumption (A) is satisfied);
2. Then they continue deliberating as before.

Note that after the first step, we have arrived at the situation of *Case* **(a)**, except that we replace $\mathcal{C}^0$ with $\widetilde{\mathcal{C}}^0$, and replace $\mathcal{M}_i$ with $\widetilde{\mathcal{M}}_i$ for all $i \in \mathcal{I}$. Thus the logic of *Case* **(a)** then applies, and yields the conclusions of Theorems 5 and 6, as before. □

Proof of Proposition 4.2. We will continue to use the notation developed in the proof of Theorems 5 and 6. We will first consider the Case (a), where the sets $\{\mathcal{M}_i\}_{i\in\mathcal{I}}$ are disjoint (so Assumption (A) is trivially satisfied). Let $\mathbf{B}^\emptyset$ be the log likelihood ratio matrix for the common prior π. Then for any $i \in \mathcal{I}$ and $t \in \mathbb{N}$, equation (B4) yields

$$\mathbf{B}_i^t \;=\; \mathbf{B}^\emptyset + \mathbf{H}_i^t + \sum_{c\in\mathcal{C}^t} \mathbf{E}^{y_c}, \quad \text{where} \quad \mathbf{H}_i^t \;:=\; \sum_{p\in\mathcal{P}_i^t} \mathbf{E}^{y_p}. \tag{B11}$$

In words, $\mathbf{H}_i^t$ is the log likelihood ratio matrix of all the evidence that i has *not* revealed at time t. If there is deliberative equilibrium at time t, then $\mathbf{B}_i^t = \mathbf{B}_j^t$ for all $i, j \in \mathcal{I}$. Substituting the expressions from (B11) into this equality, we get

$$\mathbf{B}^\emptyset + \mathbf{H}_i^t + \sum_{c\in\mathcal{C}^t} \mathbf{E}^{y_c} \;=\; \mathbf{B}^\emptyset + \mathbf{H}_j^t + \sum_{c\in\mathcal{C}^t} \mathbf{E}^{y_c},$$

and hence $\mathbf{H}_i^t = \mathbf{H}_j^t$, for all $i, j \in \mathcal{I}$. Thus, there is some matrix $\mathbf{H}$ such that

$$\mathbf{H}_i^t \;=\; \mathbf{H}, \qquad \text{for all } i \in \mathcal{I}. \tag{B12}$$

Now, since the sets $\{\mathcal{M}_i\}_{i\in\mathcal{I}}$ are disjoint, we have

$$\mathcal{M} \underset{(*)}{=} \mathcal{C}^0 \sqcup \bigsqcup_{i\in\mathcal{I}} \mathcal{M}_i \;=\; \mathcal{C}^0 \sqcup \bigsqcup_{i\in\mathcal{I}} (\mathcal{C}_i^t \sqcup \mathcal{P}_i^t) \;=\; \mathcal{C}^0 \sqcup \bigsqcup_{i\in\mathcal{I}} \mathcal{C}_i^t \sqcup \bigsqcup_{i\in\mathcal{I}} \mathcal{P}_i^t \underset{(\dagger)}{=} \mathcal{C}^t \sqcup \bigsqcup_{i\in\mathcal{I}} \mathcal{P}_i^t, \tag{B13}$$

where $(*)$ is by (2C) and $(\dagger)$ is by (2D). Thus, if $\mathbf{A}$ is as in equation (B10), then

$$\begin{aligned}\mathbf{A} &= \sum_{m\in\mathcal{M}} \mathbf{E}^{y_m} \underset{(*)}{=} \sum_{c\in\mathcal{C}^t} \mathbf{E}^{y_c} + \sum_{i\in\mathcal{I}} \sum_{p\in\mathcal{P}_i^t} \mathbf{E}^{y_p} \\ &\underset{(\diamond)}{=} \sum_{c\in\mathcal{C}^t} \mathbf{E}^{y_c} + \sum_{i\in\mathcal{I}} \mathbf{H}_i^t \underset{(\dagger)}{=} \sum_{c\in\mathcal{C}^t} \mathbf{E}^{y_c} + I\cdot\mathbf{H},\end{aligned}$$

where $(*)$ is by equation (B13), $(\diamond)$ is by the right-hand equation in equation (B11), and $(\dagger)$ is by equation (B12), with $I := |\mathcal{I}|$. Thus,

$$\sum_{c\in\mathcal{C}^t} \mathbf{E}^{y_c} = \mathbf{A} - I\cdot\mathbf{H}. \tag{B14}$$

Substituting (B12) and (B14) back into the left-hand equation in (B11), we get

$$\mathbf{B}_i^t = \mathbf{B}^{\emptyset} + \mathbf{H} + \mathbf{A} - I\cdot\mathbf{H}, = \mathbf{B}^{\emptyset} + \mathbf{A} - (I-1)\cdot\mathbf{H}, \tag{B15}$$

Now, let η be the probabilistic belief whose log likelihood ratio matrix is $\mathbf{B}^{\emptyset} - (I-1)\cdot\mathbf{H}$. Then for all $i \in \mathcal{I}$, equation (B15) tells us that $B_i^t = \alpha\cdot\eta/(\mathrm{SNC})$, as claimed.

The proof in Case (b) proceeds as in the proof of Theorems 5 and 6. □

C A game-theoretic interpretation of the model in Section 3

As emphasized just below equation (3C), the model of deliberation in Section 3 is not a game, and deliberative equilibria are not Nash equilibria. Nevertheless, Theorems 1 to 4 suggest an interesting game-theoretic reformulation of the model.

We will define a normal-form game with the following structure. The set of players is $\mathcal{I}\cup\{0\}$, where $\mathcal{I}$ is the set of deliberating agents, and 0 is "Nature". The players in $\mathcal{I}$ are all identical. Each player's strategy takes the form of a "disclosure policy", which specifies, for any possible configuration of public evidence, private evidence, and declared opinions of the other players, what evidence (if any) this player should disclose during a particular round of deliberation. During the game, each player chooses a disclosure policy, Nature determines the public evidence and private evidence of each player, and then each player receives a payoff depending on the end-result of the ensuing deliberation.

Formally, the public evidence can by summarized by an ordered pair $\mathbf{c} = (c_+, c_-) \in \mathbb{N}^2$, where c_+ is the number of positive signals that are public information, and c_- is the number of negative signals. Likewise, player i's private evidence at any stage in the deliberation can be summarized by an ordered pair $\mathbf{p}^i = (p_+^i, p_-^i) \in \mathbb{N}^2$, where p_+^i is the number of undisclosed positive signals remaining in i's private evidence, and p_-^i is the number of undisclosed negative signals. Finally, the profile of opinions of all agents can be described by an $\mathcal{I}$-tuple $\mathbf{s} = (s_i)_{i\in\mathcal{I}} \in \{-1, 0, 1\}^{\mathcal{I}}$. A *disclosure policy* is thus a function

$D : \mathbb{N}^2 \times \mathbb{N}^2 \times \{-1,0,1\}^{\mathcal{I}} \longrightarrow \{-1,0,1\}$: for any $\mathbf{c}$ and $\mathbf{p}^i$ in $\mathbb{N}^2$, and any $\mathbf{s} \in \{-1,0,1\}^{\mathcal{I}}$, we interpret $D(\mathbf{c}, \mathbf{p}^i, \mathbf{s}) = 1$ to mean that i should disclose a piece of *positive* evidence if her current deliberative situation is described by $(\mathbf{c}, \mathbf{p}^i, \mathbf{s})$ and she has the opportunity to speak. Likewise, $D(\mathbf{c}, \mathbf{p}^i, \mathbf{s}) = -1$ means that i should disclose a piece of *negative* evidence in this situation, and $D(\mathbf{c}, \mathbf{p}^i, \mathbf{s}) = 0$ means that she should disclose *no* evidence.[26]

Suppose all players in $\mathcal{I}$ choose a deliberation policy, and then Nature assigns to each of them an initial set of private evidence, along with a set of public evidence. In the parallel protocol, the ensuing deliberation is entirely determined by this information; the parallel-protocol deliberation will unfold deterministically until it reaches a *stationary state* in which all players' disclosure policies have the value 0 (so that no one discloses any further evidence). In the *serial* protocol, we must also specify an "order of speech". This can be done, for example, by labelling the players $1, 2, \ldots, |\mathcal{I}|$ and requiring them to speak cyclically in that order. Given an order of speech, the disclosure policies of the players and their initial private and public evidence once again completely determine the unfolding of the deliberation until it reaches a stationary state.

The payoff functions of all players in $\mathcal{I}$ are identical: each player receives a payoff of 1 if the deliberation reaches a stationary state that is full-disclosure equivalent. Otherwise she receives a payoff of zero.[27] The player 0 (Nature) has a constant payoff function.

The normal-form game now proceeds as follows: Nature defines an initial set of public evidence and assigns initial sets of private evidence to all players. Meanwhile, each player in $\mathcal{I}$ chooses a deliberation policy. Then deliberation unfolds deterministically, as explained two paragraphs up, and all players receive a payoff as in the previous paragraph.

Note that *a priori*, we impose no restrictions on each player's deliberation policy. For example, a player might continue to disclose evidence favourable to her opinion, even when her opinion is already endorsed by a large majority. Or she might *fail* to disclose favourable evidence, even when her side is in the minority. She might even disclose evidence which is *contrary* to her current opinion. Her disclosure strategy could depend in some complex way on the current public evidence and on the pattern of opinions of the other players. In other words, her disclosure policy might violate all of the assumptions made under the heading "Deliberative equilibrium" at the start of Section 3. But in fact, Theorems 1 to 4 tell us that she has *no interest* in violating these assumptions. These theorems tell us that, for *any* strategy chosen by Nature, the disclosure policies described at the start of Section 3 define a *Nash equilibrium* for the game described in the previous paragraph. Thus, as long as each player believes that the other players will adopt such a policy, she has no incentive to deviate from such a policy.[28]

Given this result, why did we proceed in a more roundabout way, by arguing that a combination of "good faith" and "communication costs" induces each deliberator to adopt the disclosure policy described at the start of Section 3? Because we feel that this is a

[26] We impose the feasibility constraint that $D(\mathbf{c}, \mathbf{p}^i, \mathbf{s}) \leq 0$ if $p^i_+ = 0$, while $D(\mathbf{c}, \mathbf{p}^i, \mathbf{s}) \geq 0$ if $p^i_- = 0$. It follows that $D(\mathbf{c}, \mathbf{p}^i, \mathbf{s}) = 0$ if $\mathbf{p}^i = (0,0)$.

[27] This payoff function arises because the Proposition at the start of Section 3 implies that any truth-seeking agent always prefers outcomes that are full-disclosure equivalent.

[28] We do not claim that this is the *only* Nash equilibrium for the game.

more realistic description of how people behave in actual deliberative interactions. But for those who prefer a more traditional game-theoretic analysis, one is available.

References

Acemoğlu, D., Como, G., Fagnani, F., Ozdaglar, A., 2013. Opinion fluctuations and disagreement in social networks. Mathematics of Operations Research 38 (1), 1–27.

Acemoğlu, D., Dahleh, M. A., Lobel, I., Ozdaglar, A., 2011. Bayesian learning in social networks. The Review of Economic Studies 78 (4), 1201–1236.

Acemoğlu, D., Ozdaglar, A., 2011. Opinion dynamics and learning in social networks. Dynamic Games and Applications 1 (1), 3–49.

Acemoğlu, D., Ozdaglar, A., ParandehGheibi, A., 2010. Spread of (mis) information in social networks. Games and Economic Behavior 70 (2), 194–227.

Adamowicz, W., Hanemann, M., Swait, J., Johnson, R., Layton, D., Regenwetter, M., Reimer, T., Sorkin, R., 2005. Decision strategy and structure in households: A "groups" perspective. Marketing Letters 16, 387–399.

Aumann, R. J., 1976. Agreeing to disagree. The Annals of Statistics, 1236–1239.

Austen-Smith, D., Banks, J. S., 1996. Information aggregation, rationality, and the Condorcet jury theorem. American Political Science Review 90 (1), 34–45.

Austen-Smith, D., Feddersen, T. J., 2006. Deliberation, preference uncertainty, and voting rules. American Political Science Review 100 (2), 209–217.

Bacharach, M., 1985. Some extensions of a claim of Aumann in an axiomatic model of knowledge. Journal of Economic Theory 37 (1), 167 – 190.

Bächtiger, A., Dryzek, J., Mansbridge, J., Warren, M. (Eds.), 2018. The Oxford Handbook of Deliberative Democracy. Oxford University Press, Oxford.

Bala, V., Goyal, S., 1998. Learning from neighbours. The Review of Economic Studies 65 (3), 595–621.

Bala, V., Goyal, S., 2001. Conformism and diversity under social learning. Economic theory 17 (1), 101–120.

Banerjee, A. V., 1992. A simple model of herd behavior. The Quarterly Journal of Economics 107 (3), 797–817.

Berg, S., 1997. Indirect voting systems: Banzhaf numbers, majority functions and collective competence. European Journal of Political Economy 13 (3), 557–573.

Bikhchandani, S., Hirshleifer, D., Welch, I., 1992. A theory of fads, fashion, custom, and cultural change as informational cascades. Journal of Political Economy 100 (5), 992–1026.

Bikhchandani, S., Hirshleifer, D., Welch, I., 1998. Learning from the behavior of others: Conformity, fads, and informational cascades. Journal of Economic Perspectives 12 (3), 151–170.

Castellano, C., Fortunato, S., Loreto, V., 2009. Statistical physics of social dynamics. Reviews of modern physics 81 (2), 591.

Cave, J. A., 1983. Learning to agree. Economics Letters 12 (2), 147 – 152.

Cohen, J., 1986. An epistemic conception of democracy. Ethics 97 (1), 26–38.

Coughlan, P. J., 2000. In defense of unanimous jury verdicts: Mistrials, communication, and strategic voting. American Political Science Review, 375–393.

DeGroot, M. H., 1974. Reaching a consensus. Journal of the American Statistical Association 69 (345), 118–121.

Deimen, I., Ketelaar, F., Le Quement, M. T., 2015. Consistency and communication in committees. Journal of Economic Theory 160, 24–35.

Demange, G., Van Der Straeten, K., 2017. Communicating on electoral platforms. Journal of Economic Behavior & Organization (to appear).

DeMarzo, P. M., Vayanos, D., Zwiebel, J., 2003. Persuasion bias, social influence, and unidimensional opinions. The Quarterly Journal of Economics 118 (3), 909–968.

Dickson, E. S., Hafer, C., Landa, D., 2008. Cognition and strategy: a deliberation experiment. The Journal of Politics 70 (4), 974–989.

Dickson, E. S., Hafer, C., Landa, D., 2015. Learning from debate: institutions and information. Political Science Research and Methods 3 (3), 449–472.

Dietrich, F., 2010. Bayesian group belief. Social Choice and Welfare 35 (4), 595–626.

Dietrich, F., 2019. A theory of Bayesian groups. Noûs 53 (3), 708–736.

Dietrich, F., List, C., 2016. Probabilistic opinion pooling. In: Hájek, A., Hitchcock, C. (Eds.), The Oxford Handbook of Probability and Philosophy. Oxford University Press, Ch. 25.

Dietrich, F., Spiekermann, K., 2020a. Jury theorems. In: Fricker, M., Graham, P. J., Henderson, D., Pedersen, N., Wyatt, J. (Eds.), The Routledge Handbook of Social Epistemology. Routledge (forthcoming).

Dietrich, F., Spiekermann, K., 2020b. Social epistemology. In: Knauff, M., Spohn, W. (Eds.), The Handbook of Rationality. MIT Press (forthcoming).

Doraszelski, U., Gerardi, D., Squintani, F., 2003. Communication and voting with double-sided information. Contributions in Theoretical Economics 3 (1).

Elbittar, A., Gomberg, A., Martinelli, C., Palfrey, T. R., 2016. Ignorance and bias in collective decisions. Journal of Economic Behavior & Organization (to appear).

Estlund, D., Landemore, H., 2018. The epistemic value of democratic deliberation. In: Bächtiger et al. (2018), Ch. 7, pp. 113–131.

Eyster, E., Rabin, M., 2005. Cursed equilibrium. Econometrica 73 (5), 1623–1672.

Feddersen, T., Pesendorfer, W., 1996. The swing voter's curse. American Economic Review, 408–424.

Feddersen, T., Pesendorfer, W., 1997. Voting behavior and information aggregation in elections with private information. Econometrica, 1029–1058.

Feddersen, T., Pesendorfer, W., 1998. Convicting the innocent: The inferiority of unanimous jury verdicts under strategic voting. American Political Science Review 92 (1), 23–35.

Feddersen, T., Pesendorfer, W., 1999. Elections, information aggregation, and strategic voting. Proceedings of the National Academy of Sciences 96 (19), 10572–10574.

Fishkin, J. S., Laslett, P., 2008. Debating deliberative democracy. John Wiley & Sons.

Fishman, M. J., Hagerty, K. M., 05 1990. The optimal amount of discretion to allow in disclosure. The Quarterly Journal of Economics 105 (2), 427–444.

Friedkin, N. E., Johnsen, E. C., 2011. Social influence network theory: A sociological examination of small group dynamics. Vol. 33. Cambridge University Press.

Gale, D., Kariv, S., 2003. Bayesian learning in social networks. Games and Economic Behavior 45 (2), 329–346.

Geanakoplos, J. D., Polemarchakis, H. M., 1982. We can't disagree forever. Journal of Economic Theory 28 (1), 192 – 200.

Genest, C., Zidek, J. V., 1986. Combining probability distributions: a critique and an annotated bibliography. Statist. Sci. 1, 114–148.

Gerardi, D., Yariv, L., 2007. Deliberative voting. Journal of Economic Theory 134 (1), 317–338.

Glazer, J., Rubinstein, A., 2001. Debates and decisions: On a rationale of argumentation rules. Games and Economic Behavior 36, 158–173.

Glazer, J., Rubinstein, A., 2004. On optimal rules of persuasion. Econometrica 72 (6), 1715–1736.

Goeree, J. K., Yariv, L., 2011. An experimental study of collective deliberation. Econometrica 79 (3), 893–921.

Golub, B., Jackson, M. O., 2010. Naive learning in social networks and the wisdom of crowds. American Economic Journal: Microeconomics 2 (1), 112–49.

Goodin, R. E., Spiekermann, K., 2018. An epistemic theory of democracy. Oxford University Press.

Hafer, C., Landa, D., 2007. Deliberation as self-discovery and institutions for political speech. Journal of Theoretical Politics 19 (3), 329–360.

Hagenbach, J., Koessler, F., Perez-Richet, E., 2014. Certifiable pre-play communication: Full disclosure. Econometrica 82 (3), 1093–1131.

Hartmann, S., Rafiee Rad, S., 2017. Anchoring in deliberations. Erkenntnis, 1–29.

Hartmann, S., Rafiee Rad, S., 2018. Voting, deliberation and truth. Synthese 195 (3), 1273–1293.

Houy, N., Ménager, L., 2008. Communication, consensus and order: Who wants to speak first? Journal of Economic Theory 143 (1), 140–152.

Hummel, P., 2012. Deliberation in large juries with diverse preferences. Public Choice 150 (3-4), 595–608.

Imbert, C., Boyer-Kassem, T., Chevrier, V., Bourjot, C., 2019. Improving deliberations by reducing misrepresentation effects. Episteme, 1–17.

Jackson, M. O., Tan, X., 2013. Deliberation, disclosure of information, and voting. Journal of Economic Theory 148 (1), 2–30.

Jadbabaie, A., Molavi, P., Sandroni, A., Tahbaz-Salehi, A., 2012. Non-Bayesian social learning. Games and Economic Behavior 76 (1), 210–225.

Janis, I. L., 1972. Victims of Groupthink: A Psychological Study of Foreign- Policy Decisions and Fiascoes. Houghton Mifflin Company, Boston.

Jeong, D., 2019. Using cheap talk to polarize or unify a group of decision makers. Journal of Economic Theory 180, 50–80.

Jiménez-Martínez, A., 2015. A model of belief influence in large social networks. Economic Theory 59 (1), 21–59.

Kerr, N. L., Tindale, R. S., 2004. Small group decision making and performance. Annual Review of Psychology 55, 623–656.

Klevorick, A. K., Rothschild, M., Winship, C., 1984. Information processing and jury decisionmaking. Journal of Public Economics 23 (3), 245–278.

Krasucki, P., 1996. Protocols forcing consensus. Journal of Economic Theory 70 (1), 266–272.

Ladha, K. K., 1992. The Condorcet jury theorem, free speech, and correlated votes. American Journal of Political Science, 617–634.

Ladha, K. K., 1995. Information pooling through majority-rule voting: Condorcet's jury theorem with correlated votes. Journal of Economic Behavior & Organization 26 (3), 353–372.

Landa, D., 2019. Information, knowledge, and deliberation. PS: Political Science & Politics 52 (4), 642–645.

Landa, D., Meirowitz, A., 2009. Game theory, information, and deliberative democracy. American Journal of Political Science 53 (2), 427–444.

Landemore, H., 2013. Democratic Reason: Politics, Collective Intelligence, and the Rule of the Many. Princeton Princeton University Press.

Landemore, H., Elster, J., 2012. Collective wisdom: Principles and mechanisms. Cambridge University Press.

Landemore, H., Page, S. E., 2015. Deliberation and disagreement: Problem solving, prediction, and positive dissensus. Politics, Philosophy & Economics 14 (3), 229–254.

Le Quement, M. T., 2013. Communication compatible voting rules. Theory and Decision 74 (4), 479–507.

Le Quement, M. T., Marcin, I., 2019. Communication and voting in heterogeneous committees: An experimental study. Journal of Economic Behavior & Organization (to appear).

Le Quement, M. T., Yokeeswaran, V., 2015. Subgroup deliberation and voting. Social Choice and Welfare 45 (1), 155–186.

List, C., 2018. Democratic deliberation and social choice: A review. In: Bächtiger et al. (2018), pp. 463–489.

Lobel, I., Sadler, E., 2015. Information diffusion in networks through social learning. Theoretical Economics 10 (3), 807–851.

Lobel, I., Sadler, E., 2016. Preferences, homophily, and social learning. Operations Research 64 (3), 564–584.

Lu, L., Yuan, Y. C., McLeod, P. L., 2012. Twenty-five years of hidden profiles in group decision making. Personality and Social Psychology Review 16, 54–75.

Maciejovsky, B., Budescu, D. V., 2007. Collective induction without cooperation? Learning and knowledge transfers in cooperative groups and competitive auctions. Journal of Personality and Social Psychology 92, 854–870.

Maciejovsky, B., Budescu, D. V., 2019. Too much trust in group decisions: Uncovering hidden profiles by groups and markets. (preprint; presented at the 18th Internationall Conference on Social Dilemmas in Sedona Arizona).

Mathis, J., 2011. Deliberation with evidence. American Political Science Review 105 (3), 516–529.

Mayo-Wilson, C., Zollman, K., Danks, D., 2013. Wisdom of crowds versus groupthink: learning in groups and in isolation. International Journal of Game Theory 42 (3), 695–723.

Meirowitz, A., 2007. In defense of exclusionary deliberation: communication and voting with private beliefs and values. Journal of Theoretical Politics 19 (3), 301–327.

Milgrom, P., Roberts, J., 1986. Relying on the information of interested parties. The RAND Journal of Economics, 18–32.

Mossel, E., Neeman, J., Tamuz, O., 2014a. Majority dynamics and aggregation of information in social networks. Autonomous Agents and Multi-Agent Systems 28 (3), 408–429.

Mossel, E., Sly, A., Tamuz, O., 2014b. Asymptotic learning on Bayesian social networks. Probability Theory and Related Fields 158 (1-2), 127–157.

Mossel, E., Sly, A., Tamuz, O., 2015. Strategic learning and the topology of social networks. Econometrica 83 (5), 1755–1794.

Mossel, E., Tamuz, O., 2010. Iterative maximum likelihood on networks. Advances in Applied Mathematics 45 (1), 36 – 49.

Mossel, E., Tamuz, O., 2017. Opinion exchange dynamics. Probability Surveys 14, 155–204.

Mueller-Frank, M., 2013. A general framework for rational learning in social networks. Theoretical Economics 8 (1), 1–40.

Mueller-Frank, M., 2014. Does one Bayesian make a difference? Journal of Economic Theory 154, 423–452.

Neilson, W. S., Winter, H., 2008. Votes based on protracted deliberations. Journal of Economic Behavior & Organization 67 (1), 308–321.

Ottaviani, M., Sørensen, P., 2001. Information aggregation in debate: who should speak first? Journal of Public Economics 81 (3), 393–421.

Ottaviani, M., Sørensen, P. N., 2006. Professional advice. Journal of Economic Theory 126 (1), 120–142.

Parikh, R., Krasucki, P., 1990. Communication, consensus, and knowledge. Journal of Economic Theory 52 (1), 178 – 189.

Pivato, M., 2013. Voting rules as statistical estimators. Social Choice and Welfare 40 (2), 581–630.

Pivato, M., 2017. Epistemic democracy with correlated voters. Journal of Mathematical Economics 72, 51–69.

Rivas, J., Rodríguez-Álvarez, C., 2017. Deliberation, leadership and information aggregation. The Manchester School 85 (4), 395–429.

Rosenberg, D., Solan, E., Vieille, N., 2009. Informational externalities and emergence of consensus. Games and Economic Behavior 66 (2), 979 – 994, special Section In Honor of David Gale.

Schnakenberg, K. E., 2015. Expert advice to a voting body. Journal of Economic Theory 160, 102–113.

Schnakenberg, K. E., 2017. Informational lobbying and legislative voting. American Journal of Political Science 61 (1), 129–145.

Schulte, E., 2010. Information aggregation and preference heterogeneity in committees. Theory and Decision 69 (1), 97–118.

Schwartzberg, M., 2015. Epistemic democracy and its challenges. Annual Review of Political Science 18, 187–203.

Sethi, R., Yildiz, M., 2012. Public disagreement. American Economic Journal: Microeconomics 4 (3), 57–95.

Shin, H. S., 1994. The burden of proof in a game of persuasion. Journal of Economic Theory 64 (1), 253–264.

Solomon, M., 2006. Groupthink versus the wisdom of crowds: The social epistemology of deliberation and dissent. The Southern Journal of Philosophy 44 (S1), 28–42.

Stasser, G., Stewart, D. D., Wittenbaum, G. M., 1995a. Expert roles and information exchange during discussion: The importance of knowing who knows what. Journal of Experimental Social Psychology 31 (3), 244–265.

Stasser, G., Taylor, L. A., Hanna, C., 1989. Information sampling in structured and unstructured discussions of three-and six-person groups. Journal of Personality and Social Psychology 57 (1), 67.

Stasser, G., Titus, W., 1985. Pooling of unshared information in group decision making: Biased information sampling during discussion. Journal of Personality and Social Psychology 48 (6), 1467.

Stasser, G., Titus, W., 1987. Effects of information load and percentage of shared information on the dissemination of unshared information during group discussion. Journal of Personality and Social Psychology 53 (1), 81.

Stewart, D. D., Stasser, G., 1995b. Expert role assignment and information sampling during collective recall and decision making. Journal of Personality and Social Psychology 69 (4), 619.

Szembrot, N., 2017. Are voters cursed when politicians conceal policy preferences? Public Choice 173 (1-2), 25–41.

Tamuz, O., Tessler, R. J., 2015. Majority dynamics and the retention of information. Israel Journal of Mathematics 206 (1), 483–507.

Tindale, R. S., Kluwe, K., 2015. Decision making in groups and organizations. In: Keren, G., Wu, G. (Eds.), The Wiley Blackwell Handbook of Judgment and Decision Making. John Wiley & Sons, Chichester, UK, pp. 849–874.

Van Dijk, F., Sonnemans, J., Bauw, E., 2014. Judicial error by groups and individuals. Journal of Economic Behavior & Organization 108, 224–235.

Visser, B., Swank, O. H., 2007. On committees of experts. Quarterly Journal of Economics 122 (1), 337–372.