Public Choice (2021) 189:515–531
https://doi.org/10.1007/s11127-021-00904-y

# Evidence and fully revealing deliberation with non-consequentialist jurors

**Jianan Wang**[1]

Received: 27 April 2020 / Accepted: 28 April 2021 / Published online: 20 June 2021


**Abstract**
We analyze a model of binary choice by a committee, when information is hard and pre-voting deliberation is allowed. Each member has, independently of the others, a positive probability of getting a private signal about the true state; with the remaining probability the member is uninformed. Hard information means that lying is disallowed during deliberation—informed members can reveal publicly or hide their signals, while uninformed voters have to disclose their ignorance. We allow non-consequentialist members whose thresholds for switching to the non-status-quo action vary with the number of informative signals. We show that in general, committee members will never reveal information fully during deliberation, even when we rule out partisan types who want the same action in all states. In particular, unanimity rule performs no worse than other rules.

**Keywords** Deliberation · Information revelation · Strategic voting · Collective choice · Non-consequentialist

**JEL Classification** D72 · D90

## 1 Introduction

The Condorcet Jury Theorem (1785), a milestone in collective choice theory, says that a group is more likely than an individual to choose the right one among several alternatives.[1] However, Austen-Smith and Banks (1996) note that proofs of the Condorcet Jury Theorem are statistical in nature and the key assumption is that individuals vote sincerely—in other words, an individual behaves in the same manner whether she is the sole decision-maker or one of several committee members. They and the subsequent literature showed that rational individuals should not vote sincerely (See Feddersen & Pesendorfer, 1996, 1997, 1998;

[1] See Grofman and Feld (1988) and Ladha (1991) for subsequent explorations along these lines.

✉ Jianan Wang
sufewjn@163.sufe.edu.cn

[1] School of Economics, Shanghai University of Finance and Economics, Guoding Road No. 777, Yangpu Area, Shanghai, China

Myerson, 1998; Wit, 1998; McLennan, 1998). In fact, a strategic voter should vote conditional on the other voters' signals being such that she is *pivotal*, which means that her decision will determine the final outcome for the committee. The key is that being pivotal supplies extra information about others' signals and influences the vote. However, that literature validates the conclusion of Condorcet by showing that even with strategic voters, the probability of the correct decision goes to unity as the electorate increases in size (See Feddersen & Pesendorfer, 1997).

In the literature summarized above, information is given exogenously.[2] The literature on deliberation that follows maintains the exogeneity of information but allows voters to communicate before voting. An important contribution here is Coughlan (2000), who considers the communication stage as a straw poll and shows that deliberation indeed sufficiently aggregates information under any voting rule when the committee members' preferences are known and identical. However, Austen-Smith and Feddersen (2006) point out that that result doesn't extend to situations wherein committee members' preferences exhibit heterogeneity and uncertainty. Gerardi and Yariv (2007) consider a general voting setup and show that deliberations render all veto-free rules equivalent with respect to the sequential equilibrium outcomes they generate.[3]

We study a model in which jurors (also referred to as voters) vote on two alternatives, acquit (A) or convict (C). The defendant is either guilty ($\mathcal{G}$) or innocent ($\mathcal{I}$), which is unobservable; with some probability a voter is uninformed and with the remaining probability he gets a signal that is informative about the true state of the world. We allow voters to engage in deliberation, which we model as voters simultaneously sending one public message each about their signals. We assume hard information, whose veracity can be ascertained by material evidence or logical deduction; the voting literature instead usually models deliberation as cheap talk.[4] However, it cannot be verified if a person who is silent is uninformed or concealing his information.[5] We also allow non-consequentialist preferences in addition to the standard consequentialist preferences that rely on calculating a Bayesian posterior; our voters may care not only about the decision regarding the guilt (or innocence) of the defendant, but also about the revealed profile of signals. Our objective is to answer the following question: Which voting rules sustain truthful revelation during the deliberation stage? We find that allowing hard information and a rich set of preferences (in a sense made precise subsequently) means that no voting rule induces truthful revelation in the debate.

Let us present the most closely related results in the literature and contrast them with ours. Austen-Smith and Feddersen (2006) consider a model with cheap-talk deliberation (and thus no hard information) as well as purely consequentialist voters. Their main result is that, in general, majority rule is superior to unanimity because the latter never permits *fully revealing equilibria* (FRE) in deliberation to exist, while the former may have FRE. Van Weelden (2008) extends their result on the impossibility of full information sharing

[2] Martinelli (2006) considers a situation in which voters can acquire some costly information.

[3] See Austen-Smith and Feddersen (2009) for a review of this literature.

[4] We later discuss some exceptions to that general assumption.

[5] To take a concrete example, consider a committee that is deliberating and then voting on whether a student is to be expelled for possible dishonesty. A member of the committee who had suspected the student of plagiarism can hide his observations; other members of the committee do not have access to the material leading to that inference, and even if they did, one cannot prove that the person who saw it was aware of the plagiarism.

under the unanimity rule to all voting rules when communication is sequential. Mathis (2011) introduces hard information into deliberation.[6] Assuming that every committee member is equally informed, they find the result opposite to Austen-Smith and Feddersen (2006). Their result relies crucially on that assumption.[7]

In contrast to previous work, we show that allowing verifiable information and non-consequentialist preferences leads, under any voting rule, to the existence of jurors who face no potential risk from concealing their information: such jurors can be pivotal at the communication stage only when they prefer one of the two alternatives, which means that their messages are irrelevant to the final outcome whenever they prefer the other choice.

Having described the result, we now turn to the assumptions of the model. Verifiability is quite plausible in real life; for example, a medical board expelling a doctor needs views to be backed by reports and expert testimony. Hard information has increasingly caught the attention of researchers.[8] To our knowledge, Dye (1985) is the very first to show the impact of verifiable information and provide a tractable model.[9] We consider a situation in which information, but not the lack thereof, can be proved or disproved.[10] More recently, Hahn (2011) introduces the notion of partially verifiable information into a model of sequential debate among experts who are motivated by career concerns; the main result is that self-censorship may hamper the efficiency of information aggregation.[11]

The inclusion of non-consequentialist preferences is motivated by both a theoretical need to understand the impact of incorporating such preferences, and pragmatic aspects of decision making. In real life, deliberation is more than reporting "I've received a guilty signal", but to explain the evidence in detail, including justifications, the proper way of acquiring the evidence and so on. Consequently, committee members often have more complicated attitudes towards the debate's outcome than simply calculating the posterior probability of the true state of the world and comparing it with their own thresholds. Furthermore, people evaluate evidence quite differently depending on the form and the organization of evidence.[12] Therefore, it is necessary and natural to allow committee members to show more complicated attitudes towards information revealed during debate.

Specifically, committee members prefer to share the responsibility of convicting the defendant; their thresholds for switching to a non-status-quo choice thus depend on the number of informative signals, not just on Bayesian posteriors. As an example, first consider a situation in which individuals received five guilty signals, three innocent signals, and one uninformative signal. And consider a second situation wherein individuals

[6] Two other exceptions that study deliberation with hard evidence are Bade and Rice (2007) and Schulte (2010), but they consider only majority rule. Moreover, in contrast to our work, the former studies a model of information aggregation through an election with costly information acquisition and shows that the option of communication dramatically changes the incentives to acquire information; while the latter assumes preference certainty.

[7] Mathis (2011) assumes that all jurors are equally informed so that an uninformative report is off-path in a fully revealing equilibrium. Thus, Mathis (2011) has room to specify off-path beliefs, while we do not.

[8] The model of cheap talk originates with Crawford and Sobel (1982).

[9] Seidmann and Winter (1997) consider a Sender–Receiver game wherein the Sender can certify all his information. See also Milgrom (1981) and Wolinsky (2003).

[10] Mathis (2008) generalizes Seidmann and Winter (1997) to partial certifiable settings. See also Lipman and Seppi (1995) and Forges and Koessler (2005).

[11] Jackson and Tan (2013) consider a similar setup wherein experts hold verifiable information, but the final decision is made by a distinct group of voters.

[12] Pennington and Hastie (1992) carry out experiments to show that subjects made more confident decisions when evidence is organized as a narrative rather than by legal issue; Kassin and Dunn (1997) conclude that computer-animated displays have greater impact than oral testimony on mock jurors.

received three guilty signals, one innocent signal, and five uninformative signals. Now, a juror who has standard preferences must prefer the same alternative in both situations as the posteriors are the same. However, a juror in our model may prefer Convict in the first situation but prefer Acquit in the second because she finds insufficient evidence and takes the safer or less controversial option; equivalently speaking, her effective threshold for conviction increases. Such preference is more plausible for small groups such as a jury, as in the famous film *12 Angry Men*. Also, such guilt-averse preferences are related to the psychological phenomenon of diffusion of responsibility.[13] Lack of evidence makes people more reluctant when voting against the status-quo choice. With lesser evidence for conviction, the tendency grows.[14]

Theoretically, the literature on strategic voting with non-standard preferences is relatively small. Equivalently, our model can be recast as one where voters maximize expected utility subject to a cost of guilt from convicting someone, and this cost is non-decreasing in the number of uninformative signals.[15] Ellis (2016) studies strategic voting when voters have pure common values but have ambiguity aversion preferences, i.e., they display Ellsberg-type behavior.[16] Ghosh and Tripathi (2012) consider a majority election before voters know all information determining their ranking of alternatives. The key result is that an ideologue committed to one of the two alternatives beats an idealist whose state-contingent platform implements the alternative that maximizes social welfare after all information becomes available. Bhattacharya (2013) shows that state-contingent voter preferences may often lead to failure of information aggregation.

Our paper is organized as follows. The next section first presents the model, then we proceed to present all permissible preferences. Then the next section characterizes the strategies and presents the main result. We conclude this paper after the discussion section. All proofs are collected in the Appendix.

## 2 The model

Consider a committee of voters or jurors denoted by $i \in N = \{1, 2, \dots, 2n+1\}$ that has to choose an alternative $z \in \{A, C\}$. The true state of the world $\omega \in \{\mathcal{I}, \mathcal{G}\}$ is chosen by nature and is unobservable; states are assumed to be equiprobable to save notation. For concreteness, the following interpretation may be useful—the committee can acquit ($A$) or convict ($C$) a defendant who actually may be innocent ($\mathcal{I}$) or guilty ($\mathcal{G}$).

All jurors prefer $A$ (respectively $C$) if the true state is known as $\mathcal{I}$ (respectively, $\mathcal{G}$), but the situation is complicated by two kinds of incomplete information. First, although the true state is not observable, with probability $q \in (0, 1)$ a juror privately and costlessly observes a noisy informative signal ($G$ or $I$, for guilty and innocence, respectively) of the true state, while with probability $1-q$ he receives a null signal $s = 0$. A null signal is

[13] See Manning et al. (2007).

[14] One salient example is the death penalty in Japan. The death penalty system in Japan requires three bailiffs to operate simultaneously. In fact, no single bailiff is willing to do this alone because only when they act as a group will they not know exactly which one of them put down the prisoner, providing some relief for them.

[15] We will formally explain this later in the discussion of permissible preferences.

[16] The canonical maxmin expected utility model is developed by Gilboa and Schmeidler (1989). In the jury voting context, Pan (2019) shows that pivotality alone is insufficient to determine voters' best responses for ambiguous voting games. Ryan (2021) also studies a voting model with ambiguity-averse agents but allows asymmetric penalties. They show the "Jury Paradox" raised by Feddersen and Pesendorfer (1998) persists in the presence of ambiguity.

one that is known not to have any informational content. And the informative signal is as follows: $Pr(G|\mathcal{G}) = Pr(I|\mathcal{I}) = p$, $Pr(I|\mathcal{G}) = Pr(G|\mathcal{I}) = 1 - p$ with $p \in (\frac{1}{2}, 1)$. Thus, the set of signals is $S = \{G, 0, I\}$. A *state* is a profile of signals $\mathbf{s} = (s_1, s_2, \dots, s_{2n+1})$ in the set $\mathbf{S} \equiv S^{2n+1}$, and represents all relevant information. We assume that individual information is exchangeable; thus all permutations of states are treated equally.

The second kind of incomplete information is that individual $i$'s bias $b_i$ is drawn from a finite set $B$ and known only to $i$. A bias profile $\mathbf{b}$ belongs to the set $\mathbf{B} := B^{2n+1}$, where $\mathbf{b} = (b_1, b_2, \dots, b_{2n+1})$. The bias represents different moral standards or different opinions of how many positive signals are enough to support any given alternative. Deliberation is also a process of aggregating all members' preferences; thus, we study a rich set $B$ of permissible biases, which are narrowed down later. The payoff of a voter from an alternative $z \in \{A, C\}$ when the voter has bias $b$ in state $\mathbf{s}$ is $u(z, b, \mathbf{s})$. A bias $b$ thus induces a partition $\{\mathbf{S}_b(A), \mathbf{S}_b(C)\}$ of the set of all possible states for each voter, where

$$\mathbf{S}_b(C) := \{\mathbf{s} \in \mathbf{S} \mid u(C, b, \mathbf{s}) > u(A, b, \mathbf{s})\},$$

while $\mathbf{S}_b(A) \subseteq \mathbf{S}$ is the set of states with the reverse inequality. For convenience we rule out ties.

A pair $(\mathbf{b}, \mathbf{s}) \in \mathbf{B} \times \mathbf{S}$ is a *situation*, where $\mathbf{s} = (s_1, s_2, \dots, s_{2n+1})$ is the state (signal profile).

The timeline of this game is as follows:

(1) Individuals privately learn their biases and signals, and thereafter make simultaneous public *evidence-driven* statements as below.
(2) They then vote simultaneously using a $K$-rule defined below. The committee's choice is made and members' utilities are realized.

Statements are *evidence-driven* or based on *hard information*, meaning that signals can be hidden but not falsified—uninformed voters can only report 0, or abstain if the deliberation phase is a straw poll; and informed voters can either report the information they have in hand or pretend to be uninformed voters by reporting nothing. That scenario is realistic in many situations such as an investigation or a jury debate. Investigators and jurors do not simply say "I" or "G", but present a fact and interpret it logically. It then becomes natural to model the deliberation stage as partially verifiable talk instead of pure cheap talk. While the authenticity of a report can be verified, it cannot be verified whether a committee member had relevant information. Specifically, letting $M_s \subseteq S$ denote the set of feasible messages for a voter who observed signal $s$, we impose the requirement that $M_G = \{G, 0\}$; $M_I = \{I, 0\}$; $M_0 = \{0\}$. The set of messages that could be sent in some state is $M \equiv \cup_{s \in S} M_s$, with generic element $m$. Since $M_s = \{s, 0\}$, it follows that $M = S$.

The class of voting rules we consider takes $A$ as the status quo policy, and includes unanimity ($K = 2n + 1$) and simple majority ($K = n + 1$):

**Definition 1** *K-rule*. For any integer $K \in [1, 2n + 1]$, $K \in \mathbb{Z}$, the $K$-rule chooses $A$ unless at least $K$ votes are cast for $C$.

We now proceed to the details of permissible biases in our model.

### 2.1 Permissible biases

Our goal is to compare several voting rules under a wide class of preferences influenced by evidence-driven statements. What kind of preferences do we admit? Some, such as preferring $C$ irrespective of the signal profile, are not realistic. The standard setup of biases in the deliberation literature takes the form of a cutoff probability: a juror prefers $C$ if the posterior probability of the true state being $\mathcal{G}$ exceeds that threshold. We allow committee members to exhibit more complicated attitudes towards the debate's outcome such that they not only care about the posterior probability of the true state, but also about the informativeness of the debate. Some behavioral and psychological motivations for that assumption have already been offered.

To formalize the discussion here, we introduce some more notation. The evidence-degree $e(\mathbf{s})$ of a state is the number of informed voters in that state. Denote $\mathbf{S}^r$ as the set of all possible states with $r$ informative signals, i.e., $\mathbf{S}^r = \{\mathbf{s}|e(\mathbf{s}) = r\}$. So the set $\mathbf{S}_b(A)$ induced by bias $b$ can be partitioned further into $2n + 2$ states $\mathbf{S}_b^r(A) \subset \mathbf{S}^r$ for any $0 \leq r \leq 2n + 1$, where $\mathbf{S}_b^r(A)$ is a set of states with $e(\mathbf{s}) = r$ in which $i$ would prefer $A$ to $C$. Thus, for any $r$ and any $b \in B$, we have $\mathbf{S}_b^r(C) \cup \mathbf{S}_b^r(A) = \mathbf{S}^r$. The partition set $\mathbf{S}_b^r(A)$ is unique. We first adopt the following axioms to narrow down the class.

Signals can be ordered by a binary relation; write $s \succ s'$ if $s$ is a signal that yields a higher posterior probability of $\mathcal{I}$ than $s'$. Axiom 1 means that individuals agree on which signal is more supportive of $\mathcal{I}$ and evidence is not polarizing.

**Axiom 1: Signal monotonicity** For any $s, s' \in S$ such that $s \succ s'$, any $\mathbf{s} \in S^{2n+1}$, any bias $b$, and any $i$,

$$(\mathbf{s}_{-i}, s') \in \mathbf{S}_b(A) \Rightarrow (\mathbf{s}_{-i}, s) \in \mathbf{S}_b(A);$$

similarly, $(\mathbf{s}_{-i}, s) \in \mathbf{S}_b(C) \Rightarrow (\mathbf{s}_{-i}, s') \in \mathbf{S}_b(C)$.

Note that the implied signs above cannot necessarily be reversed. Axiom 1 requires all individuals to share a common view in evaluating any specific piece of evidence. Although it imposes some limitation both psychologically and theoretically, it is natural to require all individuals to share a common view of any specific evidence; without that, a model of a committee seeking to make a group decision is not reasonable.[17]

Axiom 2 is a mild one, that no one wants to convict without evidence.

**Axiom 2: Presumption of innocence** $(0, 0, \ldots, 0) \in \mathbf{S}_b(A)$ for any $b \in B$.

Axiom 1 and Axiom 2 allow a rich set of permissible biases. In particular, we allow both consequentialist and non-consequentialist preferences. If we view consequences as the realization of actions and the true state of the world, the former care only about the consequences, but the latter also care about how many people share responsibility for it. Conviction carries a higher moral cost when very few signals are informative. In the standard literature, preferences are modeled as consequentialists, state $\mathbf{s}$ signals the true state of the

[17] See Austen-Smith and Feddersen (2006) for some psychological explanations.

world $\omega$, and two states that induce the same posterior are treated the same way. That is not the case here: for example, suppose for some bias $b$ we have $\mathbf{s}_1 = (0, \dots, 0, I, G) \in \mathbf{S}_b(C)$, yet $\mathbf{s}_2 = (0, \dots, 0, 0, 0) \in \mathbf{S}_b(A)$ by Axiom 2, although $s_1$ and $s_2$ imply the same posterior probability of the true state of the world being $\mathcal{G}$.

Axiom 3 excludes jurors who are so biased that they always vote for a single alternative regardless of the debate. The axiom merely rules out trivial cases. Notice that Axiom 2 ensures that $\mathbf{S}_b(A) \neq \emptyset$.

**Axiom 3: No partisan Types** $\forall b \in B$, $\mathbf{S}_b(A) \neq \mathbf{S}$.

We assume that every possible type of bias that is not eliminated by Axioms 1-3 is permissible in $B$. Besides, Axiom 4 is a richness assumption ensuring that every possible situation occurs with positive probability.

**Axiom 4: Full support** Every situation $(\mathbf{b}, \mathbf{s}) \in \mathbf{B} \times \mathbf{S}$ occurs with positive probability: $P(\mathbf{b}, \mathbf{s}) > 0$, *where* $B$ includes all type of permissible biases.

In Austen-Smith and Feddersen (2006), the inferiority of unanimity rule holds as long as the feasible set of biases is not a singleton, while the result under non-unanimity rule is not clear. We adopt a stronger assumption of Full Support here such that all permissible biases occur with positive probability; the advantage is that we are able to derive a clear result under all voting rules. Neither result nests the other.

***Remark*** The harshest type of bias is used to prove our result. That type of bias represents an attitude wherein a reduction in the evidence-degree magnifies the guilt from conviction, thereby increasing the threshold at which jurors switch to the non-status-quo choice. We refer to such preferences with a threshold that is non-decreasing in the evidence-degree as *guilt-averse* preferences. The harshest type is an example of such bias; it cannot be captured by assuming a constant threshold over the true state of the world: voters with harshest type of bias prefer $A$ at $\mathbf{s} = (I, 0, 0)$ but prefer $C$ at $\mathbf{s}' = (I, I, G)$, although $\mathbf{s}$ and $\mathbf{s}'$ induce the same posterior.

A utility function below shows how we can represent some non-consequentialist jurors as Bayesian jurors with multiple posteriors—which of the posteriors is used depends on the number of informative signals. Thus, a juror feels less disutility recommending punishment for the defendant when more signals are sent. Some psychological justification is provided in the introduction.

Formally, let $e(\mathbf{s}) = r$, i.e., of $2n + 1$ individuals, $r \in \{0, 1, \dots, 2n + 1\}$ of them are informed. Individual $i$ has a personal attitude $t_i$ towards two types of committee decision errors (acquitting the guilty or convicting the innocent); when he votes for conviction, he suffers a moral guilt cost of $f_i(r)$. We allow $t_i$ and $f_i(r)$ to differ across individuals. We assume that $f_i(r)$ is a non-increasing function with $f_i(0) = \frac{1}{2}$ and $f_i(2n + 1) = 0$ for all individuals. For example, take $f_i(r) = \frac{1}{2}[1 - (\frac{r}{2n+1})^2]$. Denote the final committee decision as $z \in \{A, C\}$, and denote the true state of the world as $\omega \in \{\mathcal{I}, \mathcal{G}\}$. Each individual $i$'s utility function takes the following form:

$$U_i(z,\omega|e(\mathbf{s})=r)=\begin{cases}0 & z=A,\omega=\mathcal{I}\\ -f_i(r) & z=C,\omega=\mathcal{G}\\ -t_i-f_i(r) & z=C,\omega=\mathcal{I}\\ -(1-t_i) & z=A,\omega=\mathcal{G}\end{cases}$$

Bayes rule tells us that if individual $i$ assigns a posterior probability $\beta \in [0,1]$ to the true state being $\mathcal{G}$, he would prefer conviction over acquittal if and only if $\beta \geq (t_i + f_i(r))$. In the standard literature, $f_i(r) = 0$, $\forall i, r$, so that a juror's threshold for conviction is $\beta_{min} = t_i$, which is constant when the amount of evidence $r$ varies. Here $f_i(r)$ can be positive as long as it is non-increasing in $r$, representing guilt-averse preferences.

## 3 Strategies and the main result

Having laid out the model, we now define voters' strategies and proceed to the main result. During deliberation each voter first makes an evidence-driven statement given by the function $\mu : B \times S \to M_s$, where $M_s = \{s, 0\}$ for any $s \in S$. After voters observe the debate result, which is a message profile $\mathbf{m} = (m_1, m_2, \ldots, m_{2n+1}) \in M^{2n+1} \equiv \mathbf{M}$, they cast their votes simultaneously. Each voting strategy is a function $v : B \times S \times \mathbf{M} \to \{A, C\}$ mapping the voter's private information, which is her personal bias and private signal, and every message profile into a voting decision.

In such a deliberation game, an individual must consider the case wherein either she is pivotal at the voting stage or the communication stage. Here we focus on the case in which the only binding constraint is an individual's incentive compatibility constraint at the communication stage, because the central issue we explore here is whether individuals are willing to reveal their private information truthfully.

**Definition 2** *Fully Revealing Message Strategy*. A message strategy $\mu$ is fully revealing if, for all $i \in N$, for any pair of $(b_i, s_i) \in B \times S$, $\mu(b_i, s_i) = s_i$.

Note that, conditional on all members telling the truth, the simple sincere voting strategy is undominated: each individual votes for the option yielding the largest expected payoff conditional on his or her information (see Austen-Smith and Banks 1996). Therefore, we now define a fully revealing equilibrium.

**Definition 3** *Fully Revealing Equilibrium*. A profile $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_{2n+1})$ of fully revealing message strategies and a profile $\boldsymbol{v} = (v_1, \ldots, v_{2n+1})$ of weakly undominated voting strategies constitutes a fully revealing equilibrium $(\boldsymbol{\mu}, \boldsymbol{v})$ if reporting truthfully is a Bayesian Nash equilibrium of the messaging game when voters anticipate that voting will be sincere given the deliberation.

The above therefore is a PBE of the voting game taken as a whole. Now we proceed to the main result of this paper.

**Theorem 1** *Assuming Signal Monotonicity, Presumption of Innocence, No Partisan Types, and Full Support, a fully revealing equilibrium does not exist for any K-rule.*

For any non-unanimity rule, we prove the result by showing that a juror with the harshest type of bias, i.e, he prefers $A$ only if no guilty signal $G$ is sent in a realized state, has no potential risk of misleading other jurors: *Signal Monotonicity* and *Presumption of Innocence* ensures that all jurors prefer $A$ in any state wherein no guilty signal $G$ is sent. Hence, when he prefers $A$, concealing a signal of $I$ won't mislead others to vote for $C$, while the gain of persuading others to vote for $C$ in a state where he prefers $C$ still exists. Besides, the veto power of the unanimity rule still makes a juror with extreme bias experience no loss from concealing information. Therefore, for any voting rule, there exists a type of juror such that it is beneficial for him to conceal.

Unlike Austen-Smith and Feddersen (2006), wherein jurors care only about the posterior, when jurors are non-consequentialists who want to share the responsibility for conviction, they tend to take more conservative actions and don't switch to the alternative choice when they meet a situation with less evidence revealed in a fully revealing equilibrium. Hence, only a juror with the harshest type is justified here. And under any non-unanimity rule, as he cannot mislead other harshest juror types to switch to alternative $C$ when he prefers $A$ given the realized state, the harshest type of voter is pivotal only when he prefers $C$ given the realized state, and thus concealing information is always beneficial.

Note that the nonexistence of FRE holds in spite of our axioms eliminating partisan types. We do so because partisan types by definition want the same action in all states and thus would conceal anything that would go against their preferred alternative. Moreover, the proof proceeds by demonstrating the existence of at least one partisan and one profile of reports in which that type would not want to report truthfully. The result requires those types to exist with any positive probability. Such an attitude is intuitive as it can be described as a type of guilt-averse preference as motivated earlier. However, that is not to say that other types won't conceal information; indeed we conjecture that other types might do the same. We proceed to present more discussion on that possibility in the next section.

## 4 Discussion

The harshest type of bias is essential to the proof of Theorem 1. In this section, we provide more justifications for such a bias and link the present paper to the literature.

One might argue such a type is an almost partisan type, and the result for any non-unanimity rule is expected to be more positive if we adopt a standard setup of bias (a constant cutoff as discussed before) but allow only reasonable types (i.e., we rule out partisan and almost partisan types) in a deliberation game with hard information. In other words, the veracity of the following conjecture needs to be checked.[18]

**Conjecture 1** *Consider consequentialist preferences, which are based on the posterior only. Assuming signal monotonicity, no partisan or almost partisan types, and full support of all reasonable biases, a Fully Revealing Equilibrium exists for any K-rule where* $2 \leq K \leq 2n$.

[18] We thank a referee for suggesting this.

However, even with only standard and reasonable biases being allowed, the non-existence of a fully revealing equilibrium under any non-unanimity rule still carries over in many cases. To see that, we present a counterexample below.

*Counterexample to Conjecture 1* Consider a three-person jury with standard 'guilt-neutral' preferences. For simplicity,[19] we assume that the set of reasonable biases $B$ contains only two types of biases: lenient ($\ell$) and harsh ($h$); each is drawn independently with probability $\frac{1}{2}$. A $b$-biased voter's preference is captured by a threshold for conviction $\beta_b$, i.e., he/she prefers alternative $C$ if the posterior probability of the true state being $\mathcal{G}$ exceeds $\beta_b$. Specifically, an $\ell$-biased voter has a threshold for conviction of $\beta_\ell = 0.65$ and a $h$-biased voter has a threshold of $\beta_h = 0.35$. Let the quality of the signal be $p = 0.6$, and the probability of one voter being informed be $q = 0.8$. We assume hard information as always. Voters are allowed to communicate before voting.

Notice that now an $h$-biased voter is not too biased—he is neither a partisan type nor an almost partisan type; his opinion of the relative merit of the two alternatives also is not so close to a $\ell$-biased voter's that a result similar to Coughlan (2000) applies. Hence, we call the two biases reasonable. Now we proceed to prove that there exists no fully revealing equilibria under majority rule in such a committee.

Suppose that a fully revealing equilibrium exists. A juror must consider the situation in which she is pivotal at the communication stage. Consider the case that a harsh type juror has observed an innocent signal $I$.

Under majority rule, that juror's speech can change the outcome only in two situations: $S_1$, both of the other two jurors are lenient and both observed guilty signals; $S_2$, both of the other two jurors are harsh types, one of them is uninformed, and the other observes an innocent signal. Under situation $S_1$, the harsh type juror prefers to conceal his signal so as to lead all lenient type jurors to vote for $C$, and $C$ will be the outcome. Under situation $S_2$, concealing would mislead all other harsh type jurors to vote for $C$, and $C$ will be the final outcome, while $i$ prefers $A$ at state $(I, I, 0)$.

Denote the difference in payoffs between truthfully revealing and concealing in situation $S_m$ as $\Delta\text{EU}(S_m, h, I)$, where $m \in \{1, 2\}$. Notice that

$$\begin{aligned}\Delta\text{EU}(S_1, h, I) =& [-(1-\beta_h)Pr(\omega = \mathcal{G}|S_1)] - [-\beta_h Pr(\omega = \mathcal{I}|S_1)]\\ =& \beta_h - p,\end{aligned}$$

and

$$\begin{aligned}\Delta\text{EU}(S_2, h, I) =& [-(1-\beta_h)Pr(\omega = \mathcal{G}|S_2)] - [-\beta_h Pr(\omega = \mathcal{I}|S_2)]\\ =& \beta_h - \frac{(1-p)^2}{p^2 + (1-p)^2}.\end{aligned}$$

The probability of the realized situation being $S_1$ conditional on the harsh type juror $i$ receiving an innocent signal is:

$$\begin{aligned}Pr(S_1|I, h) =& \frac{1}{4}[Pr(\mathcal{G}|I)C_2^2 q^2 p^2 + Pr(\mathcal{I}|I)C_2^2 q^2 (1-p)^2]\\ =& \frac{1}{4} q^2 p(1-p),\end{aligned}$$

[19] We say later how to relax this assumption.

and the probability of the realized situation being $S_2$ conditional on the harsh type juror $i$ receiving an innocent signal is:

$$\begin{aligned} Pr(S_2|I,h) =& \frac{1}{4}[Pr(\mathcal{G}|I)C_2^1 q(1-q)(1-p) + Pr(\mathcal{I}|I)C_2^1 q(1-q)p] \\ =& \frac{1}{2}q(1-q)[p^2+(1-p)^2]. \end{aligned}$$

It follows that the harsh type juror who receives an innocent signal is willing to tell the truth if and only if:

$$\begin{aligned} &\sum_m \Delta\text{EU}(S_m,h,I)Pr(S_m|I,h) \geq 0 \\ &\Leftrightarrow [\beta_h - p]\frac{1}{4}q^2 p(1-p) + \left[\beta_h - \frac{(1-p)^2}{p^2+(1-p)^2}\right]\frac{1}{2}q(1-q)[p^2+(1-p)^2] \geq 0 \\ &\Leftrightarrow [\beta_h - p]\frac{1}{4}q^2 p(1-p) - \frac{1}{2}q(1-q)(1-p)^2 + \beta_h\frac{1}{2}q(1-q)[p^2+(1-p)^2] \geq 0 \\ &\Leftrightarrow -0.015q^2 + 0.011q(1-q) \geq 0 \Leftrightarrow q \leq \frac{11}{26}, \end{aligned}$$

which contradicts $q = 0.8$. Thus, the harsh type juror with a signal $I$ prefers to conceal. This completes the proof. □

In summary, our counterexample shows that for even a set of reasonable biases, a fully revealing equilibrium may not exist under majority rule. The intuition is that the $h$-biased juror $i$ prefers telling the truth only when situation $S_2$ happens in the presence of an uninformed juror. Hence, when the probability of a juror being informed $q$ is too high, $i$ infers that situation $S_2$ is less likely to happen; thus, he prefers to conceal.

To see that the set $B$ in the counterexample is not the only one that works, consider a more general set of biases than just $\ell$ and $h$. Suppose that $\ell$ and $h$ each happen with probability $\frac{1}{2}(1-\varepsilon)$ and the other types happen with probabilities summing to $\varepsilon$. When $\varepsilon$ converges to 0, the most influential pivotal cases are still $S_1$ and $S_2$, as shown in the example, and the sum of the expected utilities from the rest converges to 0. Hence, the result of the example still holds.

Let us relate that result to the literature. Intuitively speaking, such a result is driven by preference uncertainty as demonstrated by Austen-Smith and Feddersen (2006): a voter decides to hide his information because of the potential gain from misleading other voters who hold different opinions of the relative merits of the alternatives.

We assume types like the harshest one in the proof of Theorem 1. However, that does not mean that such types always need to be drawn in a committee. The result requires those types to exist only with positive probability, but not probability 1 in the actual draw since voters are maximizing their ex ante payoffs, i.e., before the committee is composed. A natural question is that, what if the harshest type is not present in a realized committee? However, intuitively speaking, as long as individuals believe that the harshest type is permissible, they are aware that voters with the harshest type of bias will conceal innocent signals. Then the question becomes whether the following strategies constitute an equilibrium: only the harshest type of voters conceal innocent signals while other types truthfully reveal, and all individuals vote sincerely according to the deliberation result. If so, then in the actual draw the final outcome is still efficient and full information equivalence is achieved if the harshest type is dropped.

Unfortunately, the answer remains negative in many cases. Intuitively speaking, consider a juror who has observed an uninformative report 0. If he believes that the harshest type occurs with positive probability and jurors with such type always conceal signal $I$, then when the probability of each juror being informed $q$ is sufficiently high, he infers that the vague report is more likely to be sent by a juror with the harshest bias; his optimal voting strategy hence is to treat that uninformative report 0 as an innocent signal $I$; while in an actual draw, that message is sent by a truly uninformed reporter. Consequently, even if the information is shared fully, in an actual committee the final voting outcome will not fully aggregate information: the wisdom of the group fails.[20] Also notice that when $q = 1$, Mathis's 2011 argument will be applied to conclude that the unanimity rule is superior to majority rule, while our result suggests that neither majority rule nor unanimity rule is clearly superior to the other.

## 5 Conclusion

In this paper, we consider a collective choice problem wherein committee members can communicate prior to voting. The main concern is to arrive at a decision that aggregates committee members' preferences efficiently. Are voters willing to truthfully reveal their private information regarding the true state of the world? How do various rules compare on that metric? We obtain the negative result that neither less than unanimity nor unanimity rules can induce fully revealing equilibrium under an evidence-driven situation with possibly uninformed committee members whose preferences can be drawn from a rich set; in particular, we allow non-consequentialist preferences, where a committee member's attitude towards two feasible alternatives can depend not only on her own signals and posteriors of the true state but also the signals disclosed during deliberation.

Two main insights follow directly from our result. Despite the argument that unanimity rule gives a committee member effective veto power and thus more room to manipulate the voting procedure, unanimity rule survives in the real world in the sense that non-unanimity rules do not necessarily outperform unanimity rule. And, of course, simple majority rule is ubiquitous. Besides, as Austen-Smith and Feddersen (2006) and Mathis (2011) find polar results, the present paper suggests the need for further work to work out the exact relationship between various rules.

Secondly, we provide a very general negative result. Verifiable messages impose a higher sense of moral taste on voters, as they are allowed only to conceal instead of freely lying. Yet we come to a more negative result that information is never fully aggregated under any voting rule if voters exhibit the slightest preference heterogeneity. At first glance, that result seems quite counterintuitive. However, Dye (1985) points out a similar paradoxical result in financial auditing environments: "more detailed reporting requirements may reduce the information publicly available". That result provides some insights for policymakers. For example, the government may not want to impose a highly rigorous audit procedure on every type of firm in a very general market; the intensity and strength of a regulatory policy should be designed carefully and selected properly.

Interesting questions remain open. A natural but difficult question is to study informative but not necessarily fully revealing equilibria under non-consequentialist preferences.

[20] Supporting examples can be provided by the author.

We likewise can study cases wherein the verification of evidence is imperfect or needs time; is more room provided for individuals to operate strategically because of such delay? Those questions are left for future research.

## Appendix

***Proof of Theorem 1*** We prove the result by contradiction. Assume that a fully revealing equilibrium $(\boldsymbol{\mu}, \boldsymbol{\nu})$ exists. Notice that in any fully revealing equilibrium, sincere voting strategies imply that given $\mathbf{m} = \mathbf{s}$, it must be that $v(b, s, \mathbf{m}) = C$ if and only if $(\mathbf{s}_{-i}, s_i) \in \mathbf{S}_{b_i}(C)$ for every $i \in N$ and $b_i \in B$. In equilibrium, for every $i \in N$, every possible $(b_i, s_i) \in B \times S$, $i$'s incentive compatibility condition must hold at the communication stage:

$$\mathrm{EU}(m_i = s_i, b_i, s_i) - \mathrm{EU}(m_i = 0, b_i, s_i) \geq 0,$$

where

$$\begin{aligned}&\mathrm{EU}(m_i, b_i, s_i)\\&= \sum_{\mathbf{b}_{-i} \in B^{2n}} \sum_{\mathbf{s}_{-i} \in S^{2n}} P(\mathbf{b}_{-i}, \mathbf{s}_{-i} | b_i, s_i)[Pr(A|\mathbf{b}, \mathbf{s}, m_i)u(A, b_i, \mathbf{s}) + Pr(C|\mathbf{b}, \mathbf{s}, m_i)u(C, b_i, \mathbf{s})].\end{aligned}$$

Here, $P(\mathbf{b}_{-i}, \mathbf{s}_{-i} | b_i, s_i)$ is the probability of situation $(\mathbf{b}, \mathbf{s}) = ((\mathbf{b}_{-i}, b_i), (\mathbf{s}_{-i}, s_i))$ being obtained conditional on the pair of $i$'s bias and signal being $(b_i, s_i)$, and $Pr(z|\mathbf{b}, \mathbf{s}, m_i)$ is the probability of $z \in \{A, C\}$ being the committee's final outcome given the bias profile $\mathbf{b}$, the state profile $\mathbf{s}$, and the message profile $(m_i, \mathbf{m_{-i}}) = (m_i, \mathbf{s_{-i}})$.

Hence, $i$'s incentive compatibility condition holds when:

$$\begin{aligned}&\mathrm{EU}(m_i = s_i, b_i, s_i) - \mathrm{EU}(m_i = 0, b_i, s_i)\\&= \sum_{\mathbf{b}_{-i} \in B^{2n}} \sum_{\mathbf{s}_{-i} \in S^{2n}} P(\mathbf{b}_{-i}, \mathbf{s}_{-i} | b_i, s_i)[Pr(A|\mathbf{b}, \mathbf{s}, s_i)u(A, b_i, \mathbf{s}) + Pr(C|\mathbf{b}, \mathbf{s}, s_i)u(C, b_i, \mathbf{s})\\&\quad - Pr(A|\mathbf{b}, \mathbf{s}, 0)u(A, b_i, \mathbf{s}) - Pr(C|\mathbf{b}, \mathbf{s}, 0)u(C, b_i, \mathbf{s})] \geq 0.\end{aligned}$$

Since $Pr(C|\mathbf{b}, \mathbf{s}, m_i) = 1 - Pr(A|\mathbf{b}, \mathbf{s}, m_i)$, we simplify as follows:

$$\begin{aligned}&\mathrm{EU}(m_i = s_i, b_i, s_i) - \mathrm{EU}(m_i = 0, b_i, s_i)\\&= \sum_{\mathbf{b}_{-i} \in B^{2n}} \sum_{\mathbf{s}_{-i} \in S^{2n}} P(\mathbf{b}_{-i}, \mathbf{s}_{-i} | b_i, s_i)[Pr(A|\mathbf{b}, \mathbf{s}, s_i)u(A, b_i, \mathbf{s}) + (1 - Pr(A|\mathbf{b}, \mathbf{s}, s_i))u(C, b_i, \mathbf{s})\\&\quad - Pr(A|\mathbf{b}, \mathbf{s}, 0)u(A, b_i, \mathbf{s}) - (1 - Pr(A|\mathbf{b}, \mathbf{s}, 0))u(C, b_i, \mathbf{s})]\\&= \sum_{\mathbf{b}_{-i} \in B^{2n}} \sum_{\mathbf{s}_{-i} \in S^{2n}} P(\mathbf{b}_{-i}, \mathbf{s}_{-i} | b_i, s_i)\{[Pr(A|\mathbf{b}, \mathbf{s}, s_i) - Pr(A|\mathbf{b}, \mathbf{s}, 0)][u(A, b_i, \mathbf{s}) - u(C, b_i, \mathbf{s})]\}\\&\geq 0.\end{aligned}$$

For any voter $i \in N$ who has observed a signal $s_i$, we define the following function for simplicity:

$$\Phi_i(s_i, 0; \mathbf{b}, \mathbf{s}) \equiv [Pr(A|\mathbf{b}, \mathbf{s}, s_i) - Pr(A|\mathbf{b}, \mathbf{s}, 0)] \times [u(A, b_i, \mathbf{s}) - u(C, b_i, \mathbf{s})]. \tag{1}$$

Therefore, rewriting $i$'s incentive compatible condition, a fully revealing equilibrium exists if and only if for every $i$ and every $(b_i, s_i) \in B \times S$, the following condition holds:

$$\sum_{\mathbf{b}_{-i}\in B^{2n}} \sum_{\mathbf{s}_{-i}\in S^{2n}} P(\mathbf{b}_{-i}, \mathbf{s}_{-i}|b_i, s_i)\Phi_i(s_i, 0;\mathbf{b}, \mathbf{s}) \geq 0.$$

For any $\mathbf{b} \in \mathbf{B}$ and any $k = 0, 1, 2, \dots, r$, define the set $\mathbf{S}^{r,k}$ as the set of states where there are $k$ innocent signals among $r$ informative ones, that is:

$$\mathbf{S}^{r,k} := \{\mathbf{s} \in \mathbf{S}^r \mid \#\{i|s_i = I\} = k\}.$$

We prove the result by dividing all voting rules into two cases.

*Case 1* $2 \leq K \leq 2n + 1$.

Fix majority rule for convenience, i.e., $K = n + 1$, although the analysis is the same for any $K$-rule when $2 \leq K \leq 2n + 1$.

Consider an extremely harsh type juror $i$ with bias $\underline{b} \in B$ who prefers $C$ as long as there is at least one guilty signal $G$, i.e., $\mathbf{S}^r_{\underline{b}}(A) = \mathbf{S}^{r,r}$ for any $r$. Consider the case where $i$ has observed an innocent signal. By *Full Support*, such a situation occurs with positive probability. And we prove the result by showing that for such a voter $i$ with $(b_i, s_i) = (\underline{b}, I)$, she strictly prefers concealing her information by pretending to be an uninformed voter in the communication stage.

At any bias profile $\boldsymbol{b}$, for any signal $s_i$ observed by voter $i$, we define

$$\mathbf{Z}_i(\mathbf{b}, s_i) \equiv \{\mathbf{s_{-i}} \in S^{2n}|Pr(A|\mathbf{b}, \mathbf{s}, I) \neq Pr(A|\mathbf{b}, \mathbf{s}, 0), P(\mathbf{b}_{-i}, \mathbf{s}_{-i}|b_i, s_i) > 0\}$$

as the set of signals of other voters that have positive probability and make player $i$ pivotal.

Using $\mathbf{Z}_i(\mathbf{b}, s_i)$, the harsh type's IC condition reduces to:

$$\text{EU}(I, \underline{b}, I) - \text{EU}(0, \underline{b}, I) = \sum_{\mathbf{b}_{-i},\mathbf{s}_{-i}} P(\mathbf{b}_{-i}, \mathbf{s}_{-i}|\underline{b}, I)\Phi_i(I, 0;\mathbf{b}_{-i}, \mathbf{s}_{-i}) \geq 0, \quad (2)$$

where the index $\boldsymbol{s}_{-i}$ is summed over $\mathbf{Z}_i((\mathbf{b}_{-i}, \underline{b}), I)$.

If $i$ prefers $A$, suppose it is at state $\mathbf{s}$; by construction of $\underline{b}$ there exists no guilty signal $G$, i.e., $\mathbf{s} \in \mathbf{S}^{r,r}$. By *Presumption of Innocence*, for any $b$, $(0, 0, \dots, 0) \in \mathbf{S}_b(A)$, then by *Signal Monotonicity*, for any $b$ and any state $\mathbf{s}' \in \mathbf{S}^{1,1}$, we have $\mathbf{s}' \subset \mathbf{S}_b(A)$. Recursively, for any $b$ and any state $\mathbf{s} \in \mathbf{S}^{r,r}$, we have $\mathbf{s} \subset \mathbf{S}_b(A)$. Hence, $\mathbf{s}$ is a state where all individuals prefer $A$, hence $i$ cannot be pivotal at $\mathbf{s}$. Therefore, the harsh type can be pivotal only in a situation $(\mathbf{b}, \mathbf{s})$ where $i$ prefers $C$; so there exists at least one guilty signal $G$ and $u(A, b_i, \mathbf{s}) - u(C, b_i, \mathbf{s}) < 0$. If $i$ is pivotal at the deliberation stage, then by *Signal Monotonicity*, it must be that sending a message 0 instead of $I$ alters the majority vote and decision from $A$ to $C$. That is, $Pr(A|\mathbf{b}, \mathbf{s}, I) = 1$ and $Pr(A|\mathbf{b}, \mathbf{s}, 0) = 0$. Then (1) implies that $\Phi_i(I, 0;\mathbf{b}_{-i}, \mathbf{s}_{-i}) < 0$. And by *Full Support*, for any $\mathbf{b}_{-i} \in B^{2n}$ and $\mathbf{s}_{-i} \in \mathbf{Z}_i((\mathbf{b}_{-i}, \underline{b}), I)$, we have $P(\mathbf{b}_{-i}, \mathbf{s}_{-i}|\underline{b}, I) > 0$. Hence if the set $\mathbf{Z}_i((\mathbf{b}_{-i}, \underline{b}), I)$ is non-empty, it follows that (2) is violated.

So all that remains is to show this non-emptiness when $K = n + 1$. Consider a type of juror who has bias $b'$ such that

$$\mathbf{S}^r_{b'}(A) = \begin{cases} \mathbf{S}^{r,r} \bigcup \mathbf{S}^{r,r-1} & \text{if } 2 \leq r \leq 2n + 1 \\ \mathbf{S}^{1,1} & \text{if } r = 1, \\ \mathbf{S}^{0,0} & \text{otherwise.} \end{cases}$$

That is, a juror with bias $b'$ prefers $C$ if there is at least one guilty signal and all other signals are 0, or if there are at least two guilty signals. Notice that this type of bias $b'$ is allowed in $B$ such that it occurs in the committee with positive probability.

Consider a realized bias profile where there are $n+1$ jurors who have bias $b'$, $n$ jurors who have bias $\underline{b}$ including $i$, and the realized state is $(I, G, 0, \dots, 0)$ including $s_i = I$. By *Full Support*, this situation occurs with positive probability. Then if $i$ reports $m_i = I$, jurors with bias $b'$ will vote for $A$ and $Pr(A|\mathbf{b}, \mathbf{s}, I) = 1$. But if $i$ reports $m_i = 0$, then jurors with bias $b'$ will vote for $C$ and $Pr(A|\mathbf{b}, \mathbf{s}, 0) = 0$. This bias profile lies in $\mathbf{Z}_i((\mathbf{b}_{-i}, \underline{b}), I)$ and thereby gives non-emptiness.

Thus $i$'s incentive compatibility condition is violated when she is of type $\underline{b}$ and sees signal $s_i = I$, contradicting the existence of a fully revealing equilibrium. This proves necessity for all majority and indeed for all $K$-rules when $2 \le K \le 2n+1$.

*Case 2* $K = 1$.

That is, as long as there is at least one juror votes for $C$, then the final outcome of the committee decision process is $C$. Consider an extremely lenient type juror $j$ who prefers $C$ if and only if all the realized signals are $G$, i.e., $\mathbf{S}_{\overline{b}}(C) = \mathbf{S}^{2n+1,0}$. Consider the case where $j$ has observed a guilty signal. By *Full Support*, such a situation occurs with positive probability. And we prove the result for $K = 1$ by showing that for such voter $j$ with $(b_j, s_j) = (\overline{b}, G)$, he strictly prefers concealing his private signal in the communication stage.

Similarly, at any bias profile $\boldsymbol{b}$, for any signal $s_i$ observed by voter $i$, we define

$$\mathbf{Z}_j(\mathbf{b}, s_j) \equiv \{\mathbf{s}_{-j} \in S^{2n} | Pr(A|\mathbf{b}, \mathbf{s}, G) \neq Pr(A|\mathbf{b}, \mathbf{s}, 0), P(\mathbf{b}_{-j}, \mathbf{s}_{-j}|b_j, s_j) > 0\}$$

as the set of signals of other voters that have positive probability and make player $j$ pivotal. Using this, the lenient type's IC condition reduces to:

$$\mathrm{EU}(G, \overline{b}, I) - \mathrm{EU}(0, \overline{b}, I) = \sum_{\mathbf{b}_{-j}, \mathbf{s}_{-j}} P(\mathbf{b}_{-j}, \mathbf{s}_{-j}|\overline{b}, G)\Phi_j(G, 0; \mathbf{b}_{-j}, \mathbf{s}_{-j}) \ge 0, \tag{3}$$

where the index $\boldsymbol{s}_{-j}$ is summed over $\mathbf{Z}_j((\mathbf{b}_{-j}, \overline{b}), G)$.

Given that the other individuals are telling the truth, juror $j$ infers that if he hears a communication result showing all reports from the others are $G$, he prefers the final outcome to be $C$ and he can unilaterally ensure that the final outcome is $C$ at the voting stage regardless of his speech, since $K = 1$. Thus he is not pivotal at this case. Otherwise, he hears a communication result showing that not all the others have observed guilty signals, and he prefers the final outcome to be $A$ now. Conditional on him being pivotal at the communication stage, it must be that concealing instead of reporting $G$ will turn the final result from $C$ to $A$. That is, for all $\mathbf{b}_{-j} \in B^{2n}$, for any $\mathbf{s}_{-j} \in \mathbf{Z}_j((\mathbf{b}_{-j}, \overline{b}), G)$, $Pr(A|\mathbf{b}, \mathbf{s}, G) = 0$, $Pr(A|\mathbf{b}, \mathbf{s}, 0) = 1$, and $u(A, b_j, \mathbf{s}) - u(C, b_j, \mathbf{s}) > 0$. Therefore, $\Phi_j(G, 0; \mathbf{b}_{-j}, \mathbf{s}_{-j}) < 0$. And by *Full Support*, for any $\mathbf{b}_{-j} \in B^{2n}$ and $\mathbf{s}_{-j} \in \mathbf{Z}_j((\mathbf{b}_{-j}, \overline{b}), G)$, we have $P(\mathbf{b}_{-j}, \mathbf{s}_{-j}|\overline{b}, G) > 0$. Hence if the set $\mathbf{Z}_j((\mathbf{b}_{-i}, \overline{b}), G)$ is non-empty, it follows that (3) is violated.

To see the non-emptiness of $\mathbf{Z}_j((\mathbf{b}_{-j}, \overline{b}), G)$ when $K = 1$, consider a type of juror who has bias $b''$ such that $\mathbf{S}_{b''}(C) = \mathbf{S}^{2n+1,0} \bigcup \mathbf{S}^{2n,0}$. That is, jurors with bias $b''$ prefers $C$ if there is at most one uninformative signal and all informative signals are guilty signals. Again, $b''$ is permissible; thus it occurs with positive probability.

Consider a realized bias profile where there are $2n$ jurors who have bias $b''$, and juror $j$ has bias $\overline{b}$, and the realized state is $(G, G, \dots, G, 0)$ including $s_j = G$. By *Full support*, this situation happens with positive probability. Then if juror $j$ reports $m_j = G$, jurors with bias $b''$ will vote for $C$ and $Pr(A|\mathbf{b}, \mathbf{s}, G) = 0$. But if $j$ reports $m_j = 0$, then jurors with bias

$b''$ will vote for $A$ and $Pr(A|\mathbf{b}, \mathbf{s}, 0) = 1$. This bias profile lies in $\mathbf{Z}_j((\mathbf{b}_{-j}, \overline{b}), I)$ and thereby gives non-emptiness.

Thus $j$'s incentive compatibility condition is violated when she is of type $\overline{b}$ and sees signal $s_j = G$, contradicting the existence of a fully revealing equilibrium. This proves the result for $K = 1$ and completes the whole proof. □

**Acknowledgements** I am grateful to Sambuddha Ghosh for constant encouragement and for detailed comments on numerous drafts. I also thank Helios Herrera, Satoru Takahashi, the editor, and two anonymous referees for several useful suggestions.

**Funding** No funds was received to assist with the preparation of this manuscript.

**Data availability** Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

**Declarations**

**Conflict of interest** The author has no conflicts of interest to declare that are relevant to the content of this article.

## References

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,* 209–217.
Austen-Smith, D., & Feddersen, T. J. (2009). Information aggregation and communication in committees. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 364,* 763–769.
Bade, S., & Rice, A. (2007). Political advocacy with collective decision making. Working Paper.
Bhattacharya, S. (2013). Preference monotonicity and information aggregation in elections. *Econometrica, 81*(3), 1229–1247.
Condorcet, M. D. (1785). *Essai sur l'application de l'analyse a la probabilité des decisions rendues a la pluralité des voix*. (in French).
Coughlan, P. J. (2000). In defense of unanimous jury verdicts: Mistrials, communication, and strategic voting. *American Political Science Review, 94*(2), 375–393.
Crawford, V., & Sobel, J. (1982). Strategic information transmission. *Econometrica, 50*(6), 1431–1451.
Dye, R. (1985). Strategic accounting choice and the effect of alternative financial reporting requirements. *Journal of Accounting Research, 23*(2), 544–574.
Ellis, A. (2016). Condorcet meets Ellsberg. *Theoretical Economics, 11,* 865–895.
Feddersen, T., & Pesendorfer, W. (1996). The swing voter's curse. *American Economic Review, 86*(3), 408–424.
Feddersen, T., & Pesendorfer, W. (1997). Voting behavior and information aggregation in elections with private information. *Econometrica, 65*(5), 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.
Forges, F., & Koessler, F. (2005). Communication equilibria with partially verifiable types. *Journal of Mathematical Economics, 41*(7), 793–811.
Gerardi, D., & Yariv, L. (2007). Deliberative voting. *Journal of Economic Theory, 134*(1), 317–338.
Ghosh, S., & Tripathi, V. (2012). Ideologues beat idealists. *American Economic Journal: Microeconomics, 4*(2), 27–49.
Gilboa, I., & Schmeidler, D. (1989). Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics, 18*(2), 141–153.
Grofman, B., & Feld, S. (1988). Rousseau's general will: A Condorcetian perspective. *American Political Science Review, 82*(2), 567–576.

Hahn, V. (2011). Sequential aggregation of verifiable information. *Journal of Public Economics, 95*(11), 1447–1454.
Jackson, M., & Tan, X. (2013). Deliberation, disclosure of information, and voting. *Journal of Economic Theory, 148*(1), 2–30.
Kassin, S., & Dunn, M. (1997). Computer-animated displays and the jury: Facilitative and prejudicial effects. *Law and Human Behavior, 21*(3), 269–281.
Ladha, K. (1991). The Condorcet Jury theorem, free speech, and correlated votes. *American Journal of Political Science, 36*(3), 617–634.
Lipman, B., & Seppi, D. (1995). Robust inference in communication games with partial provability. *Journal of Economic Theory, 66*(2), 370–405.
Manning, R., Levine, M., & andCollins, A. (2007). The Kitty Genovese murder and the social psychology of helping: The parable of the 38 witnesses. *The American Psychologist, 62*(6), 555–562.
Martinelli, C. (2006). Would rational voters acquire costly information? *Journal of Economic Theory, 129*(1), 225–251.
Mathis, J. (2008). Full revelation of information in sender–receiver games of persuasion. *Journal of Economic Theory, 143*(1), 571–584.
Mathis, J. (2011). Deliberation with evidence. *American Political Science Review, 105*(3), 516–529.
McLennan, A. (1998). Consequences of the Condorcet Jury theorem for beneficial information aggregation by rational agents. *American Political Science Review, 92*(2), 413–418.
Milgrom, P. (1981). Good news and bad news: Representation theorems and applications. *The Bell Journal of Economics, 12*(2), 380–391.
Myerson, R. (1998). Extended Poisson games and the Condorcet Jury theorem. *Games and Economic Behavior, 25*(1), 111–131.
Pan, A. (2019). A note on pivotality. *Games, 10*(2), 24.
Pennington, N., & Hastie, R. (1992). Explaining the evidence: Tests of the story model for juror decision making. *Journal of Personality and Social Psychology, 62*(2), 189–206.
Ryan, M. (2021). Feddersen and Pesendorfer meet Ellsberg. *Theory and Decision, 90*, 543–577.
Schulte, E. (2010). Information aggregation and preference heterogeneity in committees. *Theory and Decision, 69*(1), 97–118.
Seidmann, D., & Winter, E. (1997). Strategic information transmission with verifiable messages. *Econometrica, 65,* 163–170.
Van Weelden, R. (2008). Deliberation rules and voting. *Quarterly Journal of Political Science, 3*(1), 83–88.
Wit, J. (1998). Rational choice and the Condorcet Jury theorem. *Games and Economic Behavior, 22*(2), 364–376.
Wolinsky, A. (2003). Information transmission when the sender's preferences are uncertain. *Games and Economic Behavior, 42*(2), 319–326.