The Value of Consensus:

Information Aggregation in Committees with Vote-Contingent Payoffs*

Yves Breitmoser† Justin Valasek‡

October 22, 2017

Abstract

Existing theoretical and experimental studies have established that unanimity is a poor decision rule for promoting information aggregation. Despite this, two-sided unanimity (consensus) is frequently used in committees making decisions on behalf of society. This paper shows that a consensus rule can facilitate truthful communication and optimal information aggregation when voters face consequences not only for the outcome of the collective decision, but also for how they personally voted. Theoretically, we show that majority rule suffers from a free-rider problem in this setting, since agents' votes are not always pivotal. Consensus mitigates free-riding since responsibility for the committee's decision is equally distributed across all agents. We test our predictions in a controlled laboratory experiment. As predicted, if consensus is required, subjects are more truthful, respond more to others' messages, and are ultimately more likely to make the optimal decision. Our work therefore provides a rationale for consensus rule in settings where committee members are held accountable, formally or informally, for their individual voting decisions.

Keywords: committees, incomplete information, decision rules, cheap talk, information aggregation, laboratory experiment

JEL Classification Codes: D71, D72, C90.

*Thanks to Marina Agranov, Inacio Bo, Agustin Casas, Tom Palfrey, Navin Kartik, Nageeb Ali, Aniol Llorente-Saguer, Dan Friedman, Andy Schotter, Guillaume Fréchette, Steffen Huck, Andrea Mattozzi, Salvatore Nunnari, Karl Schlag, Sebastian Schweighofer-Kodritsch, Peter Norman-Sorensen and Arturas Rozenas for their helpful comments and suggestions. A special thanks to Rune Midjord and Tomás Rodríguez Barraquer for the discussions and insight that helped shape the motivation for this paper. A version of this paper was previously circulated under the title "A Rationale for Unanimity." Financial support of the WZB Berlin and the DFG (project BR 4648/1 and CRC TRR 190) is greatly appreciated. Corresponding author: Justin Valasek.

†Humboldt University Berlin

‡WZB Berlin and CESifo. Contact e-mail: justin.valasek@wzb.eu

1 Introduction

Committees are a ubiquitous institution for making social decisions in the presence of uncertainty. Ideally, committees aggregate the private information of their members and thus make more informed decisions than could be made by any one individual in isolation. This intuition was first formalized by de Condorcet (1785), who showed that if all individuals hold private information that is more likely to be "right" than "wrong," and if all individuals vote according to their private information, then a sufficiently large committee that votes via a majority rule will choose the "right" option with arbitrary precision. However, in many real-world settings, committee members may face consequences not only for the outcome of the collective decision, but also for how they personally voted. For example, FDA experts may seek to avoid blame for approving a drug that proves to have severe side-effects, members of parliament may wish to signal their ideological position to the electorate, and members of criminal-trial juries may be exposed to ex-post guilt for voting to convict a defendant who is later shown to be innocent. In these cases, the assumption that individuals will vote sincerely may fail since, relative to decisions made by a single agent, decision-making by majority dilutes individual responsibility for the committee's decision, making committee members more likely to vote according to their vote-contingent biases.

In this paper, we explore the effect of vote-contingent payoffs on communication and information aggregation in committees. Additionally, we consider the impact of the decision rule on behavior. In practice, committees use a wide range of rules to reach decisions: For example, FDA committees use a majority rule when deciding whether to approve a new pharmaceutical drug for general use, while convictions in many criminal trials require the jury to unanimously vote to either convict or acquit. Based on the findings of the existing theoretical and experimental literature, however, the use of a unanimity rule to aggregate information is puzzling – existing studies have shown that unanimity is a uniquely poor decision rule for promoting information aggregation (see Feddersen and Pesendorfer, 1998, Guarnaschelli et al., 2000, Persico, 2004, Austen-Smith and Feddersen, 2006, and Bouton et al., 2017a), and at best produces the same results as majority (Coughlan, 2000, Guarnaschelli et al., 2000 and Goeree and Yariv, 2011).[1]

In contrast, our analysis suggests that the use of a two-sided unanimity (consensus) rule can have a beneficial impact when committee-members are exposed to vote-contingent payoffs—we show, theoretically and experimentally, that a consensus rule can facilitate truthful communication of private information and optimal voting behavior. By holding all members equally

[1] One important exception is a recent paper by Chan et al. (2017), who consider the impact of the decision rule on the length of the committee's deliberation process – they show that a unanimity rule can promote information aggregation by committees, since it gives more patient members control over when to terminate deliberation. In contrast, the mechanism we present here operates via the committee member's communication and voting behavior, rather than the length of deliberation.

responsible for the committee decision, a consensus rule avoids the coordination problems that occur when the committee votes via majority.

Our analysis builds on two largely independent strands of literature. On one hand, the lack of robustness of information aggregation when committee members face vote-contingent payoffs has been well-documented by a number of theoretical studies, and payoffs linked to a committee member's vote are relevant across a range of settings: committee members may face expressive or moral biases (Brennan and Buchanan, 1984, Tyran, 2004, Huck and Konrad, 2005, Feddersen et al., 2009, and Morgan and Várdy, 2012), career concerns (Visser and Swank, 2007 and Levy, 2007), and guilt or blame for supporting an incorrect decision (for an application concerning FDA committees see Friedman and Friedman, 1990, p. 208, and Midjord et al., 2017). Additionally, vote-contingent payoffs are highly relevant for committee members that face electoral pressures: e.g., individual voting records affect the reelection chances of legislators (Canes-Wrone et al., 2002) and reelection concerns affect the decisions of elected judges on state supreme courts (Hall, 1992).

On the other hand, a similarly comprehensive literature analyzes information aggregation and decision rule design when voting is preceded by open discussion. Deliberation can significantly improve a committee's ability to aggregate information in equilibrium under any decision rule (for a theoretical argument, see Coughlan, 2000 and Gerardi and Yariv, 2007; for experimental evidence, see Guarnaschelli et al., 2000, and Goeree and Yariv, 2011). Pre-vote deliberation has also been widely studied in the case where committee members have conflicting preferences over the committee decision (that is, when agents disagree about the optimal committee decision for some information set). For example, Li (2001) and Austen-Smith and Feddersen (2006) show that, theoretically, heterogenous preferences lead to the non-existence of truthful communication as an equilibrium strategy. Experimentally, however, Goeree and Yariv (2011) show that in committees with heterogeneous payoffs, deliberation results in a high degree of truthful communication and information aggregation despite theoretical predictions to the contrary.[2]

Our paper merges these two strands of the literature and shows that the impact of vote-contingent payoffs on the effectiveness of pre-vote deliberation and information aggregation is fairly drastic, implying a strong rationale for the widespread occurrence of consensus rules. That is, while the standard Condorcet model assumes that agents have preferences over the committee outcome and the revealed state, we generalize this model to include the possibility that agents also have preferences over their individual *vote*. We show that under majority, such vote-contingent payoffs can introduce a free-rider problem that eliminates equilibria where

[2]In contrast, the impact of vote-contingent payoffs on the effectiveness of pre-vote deliberation is an understudied topic – to the best of our knowledge, communication and vote-contingent payoffs in committees has only been addressed within the context of career concerns, and has focused on the optimal level of transparency. We review this literature below.

committee members truthfully communicate their private signals and vote informatively. In particular, if there exists an information set such that agents prefer to individually vote for one option, but that the committee select the other option, then there are no equilibria with truthful communication and committee-optimal voting. In contrast, under a two-sided unanimity (consensus) rule there always exists a symmetric equilibrium with truthful communication and committee-optimal information aggregation.

To clarify the intuition, assume committee members may vote for either an "expressive" option or a "nonexpressive" option.[3] Depending on the unknown state of the world, either option may be optimal. As in related models, the committee members first receive mildly informative signals about the state of the world and, prior to voting, publicly share binary messages that, either truthfully or falsely (cheap talk), indicate their signal. If the committee chooses the committee-optimal option, then all members benefit equally regardless of their vote, but members voting for the expressive option additionally obtain the expressive payoff.

Under majority, the expressive payoff invites committee members to free-ride on the majority vote by voting expressively themselves when they are not pivotal. Therefore, agents' basic objective in the cheap talk stage is to maximize the chances that the nonexpressive option is selected by a majority when it is optimal. Potential free-riders do so by sending nonexpressive messages to strengthen their co-players beliefs that voting for the nonexpressive option is optimal, while personally free-riding and voting for the expressive option. As a result, communication is strategic rather than truthful, information aggregation is compromised, and social efficiency declines. Under consensus, however, all agents must forgo their vote-contingent payoff for the committee to select the nonexpressive option. This eliminates both the possibility of free-riding and the incentive to misreport the nonexpressive option, resulting in an equilibrium where agents communicate truthfully and uniformly vote for the option that maximizes the committee's expected payoffs.

From a more general perspective, we show that homogeneous preferences over the committee outcome is not a sufficient condition for the existence of an equilibrium with truthful communication and committee-optimal information aggregation. Rather, a sufficient condition is that all agents receive *homogeneous payoffs*. Accordingly, the novel insight of our theoretical analysis is that, given homogeneous preferences over the committee outcome and vote-contingent payoffs, consensus strictly dominates majority because consensus ensures homogeneous payoffs, while the heterogeneity of payoffs under majority can result in the nonexistence of an equilibrium with truthful communication and committee-optimal information aggregation.

While the theory is suggestive, there is some ambiguity regarding the prediction that con-

[3]We consider the example of a uniform expressive bias that is independent of the committee's decision or the realized state of the world because it is the simplest realization of an vote-contingent bias, and thus represents a minimal deviation from the standard model.

sensus dominates majority. In general, there are multiple equilibria under both voting rules; for example, there always exist so-called babbling equilibria where communication is random. Additionally, behavioral biases due to lying aversion, incorrect Bayesian updating, or level-k reasoning may also interfere with the theoretical prediction. In such cases, a controlled laboratory experiment can be a useful tool to generate insights about observed behavior in an analogous choice environment.

In the second part of our paper, we present the results of a laboratory experiment that compares committee behavior under majority and consensus in the presence of expressive payoffs. In a 2×2 design, we compare the two decision rules for two different levels of expressive payoffs. Across all four treatments, the effects are highly consistent with the theory: Under majority, subjects systematically misreport their messages in a pre-vote round of binary cheap talk and, indeed, this effect is mitigated under consensus. As predicted, consensus also outperforms majority in terms of information aggregation — to the best of our knowledge, this finding constitutes the first experimental evidence of a setting where a unanimity rule is strictly preferable to majority (previous experiments in settings with pre-vote deliberation, such as Guarnaschelli et al., 2000, and Goeree and Yariv, 2011, find either no significant difference between unanimity and majority, or less information aggregation under unanimity). Thus, we conclude that vote-contingent payoffs substantially enhance the comparative efficiency of two-sided unanimity voting and may help explain its widespread use.

Finally, we analyze the behavioral mechanism underlying this result. First, we find that the private signal and co-players' messages have a similar impact on a subject's vote under consensus, but that subjects are relatively less likely to respond to co-players' messages under majority. Second, we classify agents into distinct strategy types using a finite mixture modeling approach and find that under majority rule 20–26 percent of subjects pursue a "free-rider" strategy that is biased towards falsely reporting the non-expressive option and personally voting for the expressive option. In contrast, the proportion of subjects classified as free-riders under consensus is not significantly different from zero (with point estimates at 0). In all, these findings provide strong evidence supporting the hypothesis that the higher degree of truthful communication and lower degree of "strategic" behavior under consensus has a positive causal effect on information aggregation.

Next, we review the related literature. Section 2 introduces the theoretical model, and Section 3 presents the theoretical analysis. Section 4 describes our experiment testing the predictions of the model. Section 5 presents the analysis of the experimental results and Section 6 concludes. The formal proofs, the experimental instructions and a number of robustness checks of the econometric results are provided in the Appendix.

1.1 Literature

The issue of optimal voting rules in committees was first addressed in Austen-Smith and Banks (1996), who show that for committees of finite size, information aggregation is only achieved under the decision rule that features sincere voting as an equilibrium strategy.[4] Next, Feddersen and Pesendorfer (1998) went on to show that as the size of the committee approaches infinity, all q-rules *other* than unanimity aggregate information. The reason for the failure of information aggregation under unanimity is that an agent's individual vote is only pivotal when all other agents have voted for the non-veto option, implying that it is not optimal for an agent to veto based on their private information only. This reluctance to "veto" under a unanimity rule has also been observed experimentally by Guarnaschelli et al. (2000), who find that majority rule outperforms unanimity in the standard setting. More recently, Bouton et al. (2017a,b) demonstrate that a voting mechanism that allows agents to vote for either option, *or* to veto, can outperform both majority and the standard unanimity rule.

The results referenced above, however, pertain only to settings where committees do not deliberate prior to voting. Coughlan (2000) finds that if voting is preceded by a round of cheap-talk communication, then both majority and unanimity admit equilibria with perfect information aggregation (also see Gerardi and Yariv, 2007 for a comprehensive analysis of communication and voting). Experimentally, Guarnaschelli et al. (2000) and Goeree and Yariv (2011) find that allowing subjects to either conduct a "straw poll" or communicate via chat prior to voting largely eliminates the discrepancy found between majority and unanimity absent communication.

One strand of this literature has focused on the efficacy of communication when agents have conflicting preferences over the committee outcome (see Li, 2001 and Austen-Smith and Feddersen, 2006). In this case, agents have an incentive to deviate from truthful communication in an attempt to bias the committee outcome towards their preferred option. In contrast, the mechanism we highlight here does not rely on heterogenous preferences — preferences over the committee outcome are perfectly homogeneous in our setting. Instead, agents have an incentive to misreport their signal to persuade other agents to vote for the non-expressive option, while personally free-riding and voting to obtain the expressive payoff. We show that this mechanism induces some subjects to adopt a free-riding strategy, which may explain why we find evidence for strategic communication with vote-contingent payoffs, while Goeree and Yariv (2011) find that the rate of truthful communication is not significantly affected by heterogeneous preferences.

Lastly, our paper is related to the literature on reputation payoffs in committees, which considers the Holmstrom (1999) model of career concerns applied to a committee setting. Here,

[4]For comprehensive overviews of the literature on information aggregation in voting, see Li and Suen (2009), Gerling et al. (2005), and Palfrey (2016).

reputation payoffs depend on the agent's vote relative to the aggregate voting profile, as agents seek to maximize the principal's ex-post belief that the agent is of high ability. This literature primarily considers the question of the optimal level of transparency (see Fehrler and Hughes, 2017 for a review). Visser and Swank (2007), however, consider the optimal decision rule in a setting with communication and reputation payoffs (additionally, Levy, 2007 analyzes the optimal decision rule in a setting without communication). They find that, similar to the setting analyzed in the literature on conflicting preferences, an agent's incentive to misreport their private information stems from heterogeneity in preferences over the committee outcome.[5]

In contrast, we demonstrate that with vote-contingent payoffs, committee heterogeneity can be a function of the underlying decision rule: With majority, the committee can always achieve the optimal decision even when some committee members vote for the expressive option, which results in heterogeneous payoffs and a free-rider problem. Consensus rule, however, holds all members equally responsible for the committee decision, which facilitates committee-optimal behavior by enforcing homogeneous payoffs.

2 Framework

We consider a standard Condorcet setting with pre-vote communication (cheap talk): An odd-numbered committee of N agents, $i \in \{1,2,...,N\}$ with $N \geq 3$, chooses between two options $\{R(ed), B(lue)\}$. The committee decision, denoted by $X \in \{R,B\}$, is made via a vote, where each committee member submits a vote, $v_i \in \{R,B\}$, simultaneously with no abstentions.

Additionally, there is an underlying state of the world $\omega \in \{R,B\}$. Agents do not observe the state of the world, but each has a type s_i, drawn from a finite type space S, where their type corresponds to a private signal regarding the state. Additionally, all committee members have a common prior over the state of the world, denoted $P_R = \Pr(\omega = R)$, which is uninformative ($P_R = 1/2$). Take $\mathbf{S} = S^N$ and $\mathbf{s}$ to indicate a specific profile of signals in $\mathbf{S}$. Signals are drawn independently from S according to the probability distributions p^{ω}. Each signal is partially informative—$p(s)^R \neq p(s)^B$—but no signal perfectly reveals the state of the world—$p(s)^{\omega} > 0$ for all $s \in S$ and $\omega \in \{R,B\}$. Prior to voting, agents engage in "cheap talk" communication: each agent simultaneously send a message, $m_i \in S$, that is publicly observable. After voting, a decision rule, d^v, maps the agents votes into a committee decision; that is, the decision rule is a function, $d^v : \{v_1, ... v_N\} \rightarrow X \in \{R,B\}$.

Consistent with our motivation of studying information aggregation with vote-contingent payoffs, the committee members' payoffs are a function of the committee decision, an under-

[5]Fehrler and Hughes (2017) also provide a theoretical and experimental analysis of communication and reputation payoffs; in their setting, committee members misreport the *precision* of their signal when they have a low-precision signal and communication is observed by the principle.

lying state of the world $\omega \in \{R,B\}$, *and the agent's vote* v_i. Formally, agents' preferences are represented by a vNM utility function, $u : \{X,\omega,v_i\} \to \mathbb{R}$. This formulation is quite general, allowing for vote-contingent payoffs that depend on the committee decision and the state of the world. We impose the restriction that $u(X=\omega,\omega,v_i) > u(X \neq \omega,\omega,v_i')$ for all v_i, v_i', ω, which implies that agents' payoffs from choosing the option that matches the state is greater than any vote-contingent payoff—this restriction is not strictly necessary, but implies a payoff structure analogous to the standard models of information-aggregation.

We consider the decision rules Majority and Consensus. Majority rule is defined as follows: if more than $N/2$ agents vote for B, then $X=B$; and if less than $N/2$ agents vote for B, then $X=R$. The main element of consensus rule that we seek to model is that agents coordinate prior to voting, and then *uniformly vote* for the chosen option. Therefore, to approximate a consensus rule, we adopt the following framework: After the communication stage, agents participate in a simplified deliberation stage. If deliberation reveals a sufficient level of support for B (R), then agents coordinate on voting for B (R). This framework closely follows the procedure endogenously adopted by subjects in committee games with free-form communication—Goeree and Yariv (2011) find that subjects first share their signals and then coordinate on the voting outcome.

Formally, in the deliberation stage all agents simultaneously submit an opinion, $o_i \in \{R,B\}$. We assume that if the number of opinions for B is above a threshold for agreement, then agents coordinate on voting for option B. That is, the vector of opinions then determines the voting outcome as follows:

1. If $\#(o_i = B) = N$ then $v_i = B$ *for all* i, and $X = B$.
2. If $\#(o_i = B) < N$ then $v_i = R$ *for all* i, and $X = R$.

This decision rule implies that deliberation is binding, and that a specific threshold of agreement for option R applies—if a single agent submits an opinion of R, then all agents vote for R. It is important to note, however, that the actual threshold is unimportant for our analysis. In equilibrium, all opinions will be unanimous and our results will therefore obtain for any alternative threshold of agreement. Instead, the feature of the consensus rule that is theoretically important is the fact that for a given outcome to be chosen, all agents must vote for that outcome.

We denote this game by $\Gamma = \langle P_R, S, p^{\omega}, N, d^{v} \rangle$. Agents' strategies are duples (σ,τ), where:

- $\sigma(s|s_i)$ is the probability of message s after signal $s_i \in S$,
- $\tau(s_i,m_i,\mathbf{m})$ is the probability of vote R (for Majority; for Consensus, $\tau(s_i,m_i,\mathbf{m})$ is the probability of opinion R) after signal s_i, own message m_i, and the aggregate message profile $\mathbf{m} \in \mathbf{S}$.

The timing of the game is as follows:

1. Nature draws state $\omega \in \{R, B\}$ and sends private signals $\mathbf{s} \in \mathbf{S}$.
2. Committee members observe s_i and simultaneously send messages $m_i \in S$.
3. Committee members observe $\mathbf{m}$ and simultaneously submit votes/opinions $v_i/o_i \in \{R, B\}$.
4. Votes are counted and payoffs accrue.

The equilibrium concept we consider is symmetric Perfect Bayesian Nash. By symmetry, we mean that agents with the same information sets take the same strategies.

2.1 Example

While the model introduced above is quite general, we first restrict our attention to a simplified framework that corresponds to the example we use in our laboratory experiment. Each committee member receives a private signal from a binary signal space, $s_i \in \{R, B\}$, with $\Pr(\omega = x | s_i = x)$ equal to $\alpha \in (1/2, 1)$.

Payoffs consist of a separable common-value component and vote-contingent component. For the common-value component, each agent receives a payoff of 1 if the committee chooses the decision that matches the underlying state of the world, and a payoff of zero otherwise. For the vote-contingent component, each agent has a simple voting bias and receives an payoff of $K < 1$ conditional on voting for option R. This voting bias is unconditional on the state of the world and the decision of the committee—such an vote-contingent payoff is commonly referred to as an expressive bias. To summarize, terminal payoffs are a function of outcome $X \in \{R, B\}$, state $\omega \in \{R, B\}$, and the own vote $v_i \in \{R, B\}$ and are equal to:

$$u(X, \omega, v_i) = \begin{cases} 0, & \text{if } X \neq \omega \text{ and } v_i = B \\ 1, & \text{if } X = \omega \text{ and } v_i = B \\ K, & \text{if } X \neq \omega \text{ and } v_i = R \\ 1+K, & \text{if } X = \omega \text{ and } v_i = R \end{cases}$$

We denote the aggregate number of signals of B by $S^{\#}$ ($S^{\#} = \sum_i I_{s_i=B}$ where $I_{s_i=B}$ takes a value of 1 if the argument, $s_i = B$, is satisfied and 0 otherwise), and the aggregate number of messages of B by $M^{\#}$ ($M^{\#} = \sum_i I_{m_i=B}$). We denote this simplified game by $\hat{\Gamma} = \langle K, P_R, \alpha, N, d^v \rangle$, and define $\sigma(s_i)$, $\tau(s_i, m_i, M^{\#})$ as the probability of sending message R and vote R, respectively. We also note that in this simplified setting, the equilibria highlighted below are also equilibria in the model with one-sided unanimity and a status quo of R.

3 Truthful Communication and Information Aggregation

In this section, we present a partial characterization of the equilibria of the simplified model with binary signals under Majority and Consensus, followed by the general results. Our aim here is not to be exhaustive—games with communication generally admit a multitude of equilibria—rather we focus on characterizing the conditions under which the model supports equilibria with truthful communication and optimal voting outcomes.

Definition 1 (Committee-Optimal Information Aggregation)**.** *We define* Optimal Information Aggregation *as replicating the decision taken by a single decision-maker who maximizes aggregate expected payoffs of the committee,* $\sum^N E[\pi_i]$, *and has access to the complete profile of signals,* $\mathbf{s}$.

Note that there is a distinction between optimal information aggregation and perfect information aggregation: perfect information aggregation implies that the committee selects the option with the highest number of signals since this option is most likely to coincide with the state of the world; (committee) optimal information aggregation, however, takes into account the vote-contingent payoff, K, which implies that option R may be optimal even when $\omega = B$ is more likely. We acknowledge that in cases where a committee takes a decision on behalf of society, the appropriate social welfare function may be perfect information aggregation since the committee members' vote-contingent payoffs may be small relative to the social impact of choosing the more likely option. Therefore, while we focus on committee-optimal information aggregation in the theoretical analysis, we will consider both perfect and optimal information aggregation when interpreting our experimental results.

3.1 Voting Stage

We begin by characterizing the equilibria in the voting stage given truthful messaging. Note that given $m_i = s_i$ for all i, all agents are at the same information set, which implies that the symmetric equilibrium prescribes a uniform strategy for all agents in the voting stage. Accordingly, we simplify the notation of $\tau(s_i, m_i, M^{\#})$ to $\tau(M^{\#})$ for this subsection.

First, we state a well-known result in games of information aggregation through voting, which specifies that all agents voting R is an equilibrium strategy under both majority and consensus for any profile of messages. This equilibrium is commonly referred to as the "babbling equilibrium" (we refer to equilibria other than the babbling equilibrium as "non-babbling").

Lemma 1 (Babbling Equilibrium)**.** *It is a symmetric equilibrium strategy for all agents to set* $\tau(M^{\#}) = 1$ *for all* $M^{\#}$*, under both Majority and Consensus.*

The rationale for Lemma 1 is that when $\tau(M^{\#}) = 1$ for all $i \neq j$ then i's vote/opinion cannot be pivotal, which implies that $v_i/o_i = R$ is a best response. Note that, for $K > 0$ there is no babbling equilibrium with $\tau(M^{\#}) = 0$, since if i's vote is not pivotal, then i has a best response of voting for R and receiving the expressive payoff.

Next, we introduce some additional notation. Take $piv(x) = \frac{(N-1)!}{[(N-1)/2!]^2} x^{(N-1)/2}(1-x)^{(N-1)/2}$ for $x \leq 0.5$, which is equal to the probability that i's vote is pivotal under Majority when all other agents vote R with probability x. The following values will be helpful in the analysis:

$$\bar{S} = \min\{S^{\#} | \Pr(\omega = B|S^{\#}) - \Pr(\omega = R|S^{\#}) \geq K\},$$
$$S^{piv} = \min\{S^{\#} | piv(0.5)[\Pr(\omega = B|S^{\#}) - \Pr(\omega = R|S^{\#})] \geq K\}.$$

Optimal information aggregation prescribes that the committee play strategies such that they select option R when $S^{\#}$ is smaller than $\bar{S}$, and option B when $S^{\#}$ is greater or equal to $\bar{S}$. Also, note that if $S^{\#}$ is smaller than $\bar{S}$, then voting B is a dominated strategy, since i maximizes their expected payoff by voting R even when they are pivotal.

The following lemma characterizes all non-babbling equilibria of the voting stage given truthful messaging and a majority rule.

Lemma 2 (Voting Stage: Majority). *Given truthful messaging ($\mathbf{m} = \mathbf{s}$) and $K > 0$, consider the voting stage following $M^{\#}$ messages for B:*

- *If $M^{\#} < S^{piv}$, then the unique symmetric equilibrium strategy is to vote R with probability $\tau^*(M^{\#}) = 1$.*
- *If $M^{\#} \geq S^{piv} (\geq \bar{S})$, then the set of non-babbling symmetric equilibrium strategies is equal to $\{\tau^*(M^{\#}), (1 - \tau^*(M^{\#}))\}$ with $\tau^*(M^{\#}) \in (0, 0.5)$.*

We focus on the equilibrium $\tau^*(M^{\#})$, since $(1 - \tau^*(M^{\#}))$ is unstable. This non-babbling symmetric voting strategy is the mixed strategy that equalizes the benefit of choosing B over R multiplied by the probability of being pivotal, and the forgone expressive payoff. Rearranging the equilibrium condition gives:

$$piv(\tau^*(M^{\#})) = \frac{K}{\Pr(\omega = B|S^{\#}) - \Pr(\omega = R|S^{\#})}. \tag{1}$$

This equation also shows that $\tau^*(M^{\#})$ is increasing in $M^{\#}$, since $\Pr(\omega = B|S^{\#}) - \Pr(\omega = R|S^{\#})$ is increasing in $S^{\#}$ and $piv(\tau^*(M^{\#}))$ is increasing in $\tau^*(M^{\#})$ (since $\tau^*(M^{\#}) \leq 0.5$).

More generally, Lemma 2 shows that Majority introduces a coordination/free-rider problem in the voting stage. Given $M^{\#} \geq (N+1)/2$, all agents would prefer that the committee select option B. Under Majority, however, only a strict subset of the committee members are required to forgo the expressive payoff and vote for B for the committee to select B. Accordingly, each committee member would prefer that a majority vote for B, but to individually belong to the minority of agents that vote for R. This incentive to free-ride implies that there is no symmetric strategy where all agents vote for B, and that the only non-babbling symmetric equilibrium involves a mixed strategy in the voting stage (for $M^{\#} \geq S^{piv}$). This mixed strategy equilibrium introduces an aggregate bias toward option R, since there is a positive probability that the committee will select R even when $M^{\#} \geq S^{piv}$.[6]

In contrast, the following lemma characterizes the non-babbling equilibrium of the voting stage under Consensus.

Lemma 3 (Deliberation Stage: Consensus)**.** *Given truthful messaging (*$\mathbf{m} = \mathbf{s}$*) and* $K > 0$*:*

- *If* $M^{\#} < \bar{S}$*, then* $\tau(M^{\#}) = 1$ *is the unique symmetric equilibrium strategy.*
- *If* $M^{\#} \geq \bar{S}$*, then* $\tau(M^{\#}) = 0$ *is the unique non-babbling symmetric equilibrium strategy.*

That is, under Consensus, the unique non-babbling equilibrium optimally aggregates information.[7] The reason for this difference between Consensus and Majority is that under Consensus, the committee can only select option B if all committee members forgo the expressive payoff of K. Due to this the uniform enforcement of responsibility, the free-rider problem that occurs under Majority is mitigated, and optimality in the voting stage is restored.

3.2 Truthful Messaging

We now characterize, conditional on the equilibrium strategy in the voting stage, when truthful messaging is supported in equilibrium. We begin with the babbling equilibrium.

Lemma 4 (Truthful Messaging under Babbling)**.** *Given* $\tau(M^{\#}) = 1$ *for all* $M^{\#}$*, truthful messaging (*$\mathbf{m} = \mathbf{s}$*) is an equilibrium under both Majority and Consensus.*

That is, given babbling in the voting stage, agents are indifferent between sending a message of R and a message of B. Therefore, any messaging strategy is an equilibrium.

Take $\bar{K}$ to be equal to the following value:

$$\bar{K} = \left[\Pr\left(\omega = B | S^{\#} = \tfrac{N+1}{2} \right) - \Pr\left(\omega = R | S^{\#} = \tfrac{N+1}{2} \right) \right]$$

Note that for $K = \bar{K}$, the corresponding $\bar{S}$ is equal to $(N+1)/2$, implying that $\bar{S} > (N+1)/2$ for $K > \bar{K}$.

[6]For some K, there exists a plausible asymmetric equilibrium where agents with $m_i = R$ set $\tau_i(M^{\#}) = 1$ for all $M^{\#}$. In this case, a non-babbling equilibrium exists where agents with $m_i = B$ set $\tau_i((N+1)/2) = 0$ and $\tau_i(M^{\#}) = \tau^{**}(M^{\#})$ with $\tau^{**}(M^{\#}) \in (0, 0.5)$ for $M^{\#} > (N+1)/2$. As we show in the Supplementary Appendix, however, this equilibrium only exists for small values of K.

[7]There is no mixed-strategy equilibrium since, given a positive probability of being pivotal and $M^{\#} \geq \bar{S}$, i will strictly prefer voting B.

Also, take $K' = piv(0.5)\bar{K}$. The following proposition gives a necessary condition for truthful messaging and the non-babbling voting strategy to be an equilibrium under Majority.

Proposition 1 (Truthful Messaging under Majority)**.** *Given that agents play the non-babbling equilibrium in the voting stage, truthful messaging ($\mathbf{m} = \mathbf{s}$) is an equilibrium only if $K \leq K'$.*

Proposition 1 shows that the coordination/free-rider problem in the voting stage has a knock-on effect on the messaging stage, and that given K sufficiently high, there is no equilibrium with truthful messaging and non-babbling voting.

An intuition for this result is as follows: independently of their own vote, each agent would prefer that the committee select option B when there are more signals for B than R (information aggregation). However, as detailed in Lemma 2, under truthful communication the expressive payoff biases the committee's decision toward R, since agents play mixed strategies in the voting game given $M^{\#} \geq S^{piv}$. This bias is decreasing in $M^{\#}$ since the higher the likelihood that the state is equal to B, the higher the probability that agents vote for B. Therefore, for low levels of K ($K \leq K'$) an agent with signal $s_i = R$ faces the following tradeoff when considering a deviation from truthful messaging: (1) for $S^{\#} > (N+1)/2$, deviating to $\sigma(R) = 0$ will increase the probability that the committee 'correctly' selects option B; (2) for $S^{\#} = \lfloor (N-1)/2 \rfloor$, deviating to $\sigma(R) = 0$ results in a positive probability that the committee incorrectly selects option B since given $S = \lfloor (N-1)/2 \rfloor$ ($M^{\#} = \lfloor (N-1)/2 \rfloor$) all committee members other than i will vote B with positive probability. Given this tradeoff, truthful messaging can still be an equilibrium strategy for K low enough, since the benefit of deviating to $\sigma(R) = 0$ and decreasing the bias of the committee (1) can be offset by the cost of incorrectly aggregating information when i's information is pivotal (2).

However, by Lemma 2, if $K > K'$ it is not an equilibrium strategy for agents to vote for B with positive probability when $M^{\#} = \lfloor (N-1)/2 \rfloor$. Therefore, the cost of deviating from truthful messaging (2) disappears, which implies that when all other agents play truthfully, deviating to $\sigma(R) = 0$ is a best response. Additionally, an agent with a signal of R who deviates from truthful reporting has a best response to vote R in the voting sub-game given that other agents now play $\tau^*(S^{\#}+1) < \tau^*(S^{\#})$. This shows that truthful communication cannot be an equilibrium under Majority when $K > K'$, since agents have a best response to deviate to falsely reporting B while personally voting R (the "free-riding" strategy).[8]

In contrast to Majority, the following proposition shows that truthful messaging remains an equilibrium for all voting strategies under Consensus.

Proposition 2 (Truthful Messaging under Consensus)**.** *Under Consensus, truthful messaging ($\mathbf{m} = \mathbf{s}$) and committee-optimal voting ($\tau(M^{\#}) = 1$ for $M^{\#} < \bar{S}$; $\tau(M^{\#}) = 0$ for $M^{\#} \geq \bar{S}$) is an equilibrium of any game $\hat{\Gamma}$.*

[8] By the same logic, with the asymmetric voting strategy where agents with $m_i = R$ set $\tau_i(M^{\#}) = 1$ for all $M^{\#}$ (see Footnote 6), truthful messaging is not an equilibrium for $K > \bar{K}$.

The existence of an equilibrium with optimal information aggregation under Consensus follows from Lemma 3: given that agents efficiently aggregate information in the voting stage, agents have no incentive to deviate from truthful communication.

Independently, both Proposition 1 or Proposition 2 relate to results in the existing literature and thus concur with the general understanding of information aggregation. As discussed in the introduction, it is well-known that vote-contingent payoffs can cause information aggregation to fail. Additionally, Proposition 2 follows from the same reasoning behind Theorem 1 in Austen-Smith and Feddersen (2006), who analyze communication in a setting with heterogenous preferences over the committee outcome (also see Gerardi and Yariv, 2007).

However, Proposition 1 demonstrates that homogeneous preferences over the committee outcome, conditional on the aggregate profile of signals, is not a sufficient condition for the existence of an equilibrium with truthful communication and committee-optimal information aggregation. Instead, in a more general setting that allows for vote-contingent payoffs, the sufficient condition requires that agent receive homogeneous *payoffs*. Accordingly, the novel insight provided by the combination of Propositions 1 and 2 is that the committee's decision rule is relevant for efficiency, even in settings with communication and homogeneous preferences over the committee outcome, since *the homogeneity of payoffs is a function of the decision rule*.

Propositions 1 and 2 also allow us to compare the Pareto optimal equilibria under Consensus and Majority. Jointly, they show that when committee members are exposed to vote-contingent payoffs that condition on their individual vote, then the Pareto optimal outcome under Consensus always weakly dominates the Pareto optimal outcome under Majority in terms of information aggregation, and Consensus strictly dominates Majority for vote-contingent payoffs of intermediate size.

To be more precise, take K'' to be defined as follows:

$$K'' = \left[\Pr\left(\omega = B | S^{\#} = N\right) - \Pr\left(\omega = R | S^{\#} = N\right)\right]$$

When K is large enough so that $\bar{S} > N$, i.e., for all $K > K''$, then the unique equilibrium is for the committee to always select R under both Majority and Consensus. Fixing all parameters of the game other than K (N, P_R, α), take the following set of games:

$$\{\hat{\Gamma}\}^K = \{\hat{\Gamma} | K \in (K', K'']\}.$$

This definition allows us to define a set of games where Consensus strictly dominates Majority in terms of information aggregation.

Corollary 1 (Summary)**.** *Under Majority, for any game $\hat{\Gamma} \in \{\hat{\Gamma}\}^K$, there are no symmetric equilibria with truthful communication and informative voting, or any symmetric equilibria*

that optimally aggregate information.
Under Consensus, for any game $\hat{\Gamma} \in \{\hat{\Gamma}\}^K$ there exists a symmetric equilibrium with truthful communication that optimally aggregates information.

Corollary 1 follows from the fact that the committee can only efficiently aggregate information when playing symmetric strategies if (1) all agents vote their signal, or (2) agents play voting strategies that perfectly condition on $S^{\#}$. When K is large, (1) is precluded by Lemma 5 (stated and proved in the Appendix), and (2) is precluded by Proposition 1.[9]

3.3 Generalized results and discussion

As shown in the following theorem, the findings of Corollary 1 also apply the broader setting that admits a more general set of vote-contingent payoffs that may condition on realized state of the world, and/or the committee outcome.

Theorem 1 (Existence under Consensus)**.** *For any game Γ and utility function $u : \{X, v_i, \omega\} \to \mathbb{R}$, there exists a symmetric equilibrium under Consensus with truthful communication that optimally aggregates information.*

The intuition behind this generalized result is straightforward. Take $\mathbf{S}^R \subset \mathbf{S}$ to be the set of signals under which i weakly prefers outcome R and $\mathbf{S}^B \subset \mathbf{S}$ to be the set of signals under which i strictly prefers outcome R *given* $v_i = X$. That is:

$$\mathbf{S}^R = \{\mathbf{s} | E(u(X = R, v_i = R, \omega)|\mathbf{s}) \geq E(u(X = B, v_i = B, \omega)|\mathbf{s}),$$
$$\mathbf{S}^B = \{\mathbf{s} | E(u(X = B, v_i = B, \omega)|\mathbf{s}) > E(u(X = R, v_i = R, \omega)|\mathbf{s}).$$

By definition, $\mathbf{S}^R \cup \mathbf{S}^B = \mathbf{S}$ and $\mathbf{S}^R \cap \mathbf{S}^B = \emptyset$.

Under a Consensus rule, the outcome set is constrained to the subset of outcomes where $X = v_i$; effectively, $X = x'$ implies that $v_i = x'$ for all i. Therefore, given truthful communication, $(m_i = s_i)$, setting $o_i = R$ if $\mathbf{m} \in \mathbf{S}^R$ and $o_i = B$ if $\mathbf{m} \in \mathbf{S}^B$ in the deliberation stage is a best reply since agents have homogenous payoffs in all possible outcomes. Additionally, given that agents play the strategies $o_i = R$ if $\mathbf{m} \in \mathbf{S}^R$ and $o_i = B$ if $\mathbf{m} \in \mathbf{S}^B$, truthful communication results in the outcome that $X = R$ if and only if $\mathbf{s} \in \mathbf{S}^R$. This implies that, conditional on the the realized set of signals, any deviation from truthful communication will result in either the same outcome, or that $X \neq X'$ for some $\mathbf{s}' \in \mathbf{S}^{X'}$, which is a strictly worse outcome and thus demonstrates the result.

[9]The above analysis has focused on characterizing the existence/non-existence of equilibria with truthful communication and optimal information aggregation. In the Appendix we provide a broader characterization of equilibria for the specific case we analyze in our experiment (including asymmetric equilibria that consider the case where agents condition their vote on their signal and message).

It is also straightforward to see the possibility of non-existence of equilibria with truthful communication and optimal information aggregation under Majority in the generalized setting:

Corollary 2 (Non-existence under Majority)**.** *If there exists a vector of signals,* $\mathbf{s}'$*, and a committee outcome,* X'*, such that* $E(u(X',v_i,\omega)|\mathbf{s}') > E(u(X \neq X',v_i,\omega)|\mathbf{s}')$ *and* $E(u(X',v_i \neq X',\omega)|\mathbf{s}') > E(u(X',v_i = X',\omega)|\mathbf{s}')$*, then there are no symmetric equilibria with truthful communication and committee-optimal voting outcomes under Majority.*

That is, given vote-contingent payoffs where a vector $\mathbf{s}'$ exists such that i prefers that the committee selects outcome X', but individually vote for outcome $X \neq X'$, then vote-contingent payoffs introduce a coordination problem due to the fact that majority rule admits the outcome, $\{X', v_i \neq X'\}$. In this case, as illustrated in the simplified setting analyzed above (Lemma 2), there is are no symmetric equilibria with truthful communication and committee-optimal voting outcomes under Majority.

The findings of Theorem 1 and Corollary 2 are particularly striking in the context of the literature on decision rules and information aggregation in committees—vote-contingent payoffs precisely reverse the findings in the existing theoretical literature on information aggregation in committees that, without vote-contingent payoffs, voting outcomes are weakly optimal under Majority (Feddersen and Pesendorfer, 1998; Austen-Smith and Feddersen, 2006; Bouton et al., 2017a). To be clear, our results do not imply that Consensus dominates Majority in all situations—our model considers a decision-situation that is specifically designed to isolate the role of vote-contingent payoffs. However, it does suggest that uniformly enforcing responsibility for the committee's decision may mitigate a voting bias that occurs in many real-world settings where committee members can be held accountable, formally or informally, for their individual voting decisions.

Additionally, while Theorem 1 and Corollary 2 are suggestive, they do not provide a fully satisfactory answer as to which decision rule is optimal in a setting with vote-contingent payoffs. For example, both decision rules admit multiple equilibria, even under the assumption of symmetric strategies, leading to ambiguity regarding the theoretical predictions. Moreover, there are well-documented behavioral factors that may impact the theoretical predictions: a single agent voting naively is enough to destroy the efficiency of information aggregation under unanimity rules, and a preference for truth-telling (lying aversion) may mitigate the negative predictions for Majority. Additionally, there are more subtle ambiguities: if we allow for asymmetric strategies, note that if $(N-1)/2$ agents play a fixed strategy $\tau(M^{\#}) = 1$ for any $M^{\#}$, then the remaining $(N+1)/2$ agents face an effective decision rule of unanimity, allowing for optimal information aggregation.

This indeterminacy of the theoretical predictions leave scope for further exploration. In such a setting, where the theoretical predictions are suggestive but not definitive, laboratory experiments can be a useful tool to generate insights on observed behavior in an analogous choice environment. In the following section, we explain how our experimental design may help us to detail the differences in communication and voting under Majority and Consensus, and to explore the mechanisms behind the observed differences in aggregate behavior under the two decision rules.

4 Experiment

4.1 Experimental Design

The experiment closely implements our model of voting with expressive payoffs, using a 2×2 design with "High" and "Low" expressive payoffs under Majority and Consensus. The experimental implementation closely follows the related experiments of Guarnaschelli et al. (2000) and Goeree and Yariv (2011). In particular, we use neutral language, communicate probabilities and signals to subjects using balls drawn from urns, and provide feedback about the actual state of world and composition of payoffs after each round. A detailed description follows and a translation of the instructions and a screenshot are provided as supplementary material. The experiments were conducted at the WZB/TU experimental laboratory in Berlin in May, June and November of 2016. Subjects were recruited using ORSEE (Greiner, 2015) and the experiment was programed in Z-Tree (Fischbacher, 2007).

The four treatments are summarized in Table 1. The sum of the expressive payoff K, which committee members get after voting R, and social payoff P_c, which committee members get after voting in line with the state of the world, is always equal to to 50 points. In the treatments with "low" expressive payoffs, we set $K = 10$ and $P_c = 40$, and in the treatments with "high" expressive payoffs, we set $K = 15$ and $P_c = 35$. The payoffs were calibrated such that $K > \bar{K}$ in both the Low and High treatments, and we discuss the theoretical predictions in detail in the following subsection. In both cases, we conduct sessions with both consensus and majority voting. The precision of each subject's signal was constant across treatments and equal to $\alpha = 0.6$. Subjects were paid according to the sum of points accumulated across all 50 games, and one point corresponded to one euro cent in all treatments. The experiment lasted between 75 and 105 minutes and subjects earned between 19 and 22 Euros on average across sessions.

Table 1: Overview of experimental treatments

Label	Decision rule	P_c	K	#Subjects	#Sessions	#Games
Majority-Low	Majority	40	10	48	4	50
Majority-High	Majority	35	15	45	4	50
Consensus-Low	Consensus	40	10	45	4	50
Consensus-High	Consensus	35	15	48	4	50

For each treatment, we ran four sessions with either 9 (two sessions) or 12 (fourteen sessions) participants. In all cases, two sessions were run simultaneously to increase anonymity. Upon arrival at the laboratory, subjects were seated randomly. An experimental assistant then handed out printed versions of the instructions and read the instructions out loud. Subsequently, subjects filled in a computerized control questionnaire verifying their understanding of the instructions, and the experiment did not start until all subjects had answered all questions correctly. The subjects then played 50 voting games in committees of size three ($N = 3$), with random rematching after each game (see Figure 3 in Appendix B for a composite screenshot). After each game, the subjects received feedback on payoffs, aggregate behavior and the aggregate signal profile. Under Majority, the timing of each round was as follows.

Majority Observe private signal $\{R,B\}$. Send a public message to their group $\{R,B\}$. Observe message profile. Submit vote $\{R,B\}$. Observe state, votes, outcome, and payoffs for this game.

Under Consensus, the timing of each round was identical to Majority aside from the voting stage. Subjects were given three chances to reach a unanimous decision, after which all subjects were assigned a default vote of R. We choose to operationalize consensus rule in this manner to replicate a two-sided unanimity as closely as possible. Moreover, we choose a conservative default of R to ensure that the frequency of committee outcomes equal to the non-expressive option under Consensus is not driven by the default option.

Consensus The only difference to Majority is in the voting stage: Submit vote $\{R,B\}$. If vote is unanimous proceed to Outcome Stage. Otherwise, again submit a vote $\{R,B\}$. If vote is unanimous proceed to Outcome Stage. Otherwise, submit a final vote $\{R,B\}$. If vote is not unanimous, all subjects are assigned the vote $v_i = R$. Proceed to Outcome stage.

Additionally, the format of allowing for multiple straw polls is similar to the procedure used in many jury deliberations. It is possible, however, that subjects use the multiple rounds of voting as additional communication. We address this concern in a robustness section in the Appendix. Upon completion of the experiment, subjects left the laboratory and were paid individually in a separate room by an experimental assistant.

4.2 Theoretical Predictions and Experimental Research Questions

Corollary 1 shows that for a wide range of expressive payoffs, truthful communication is ruled out in non-babbling equilibria under Majority voting.[10] The expressive payoffs induced in the

[10]To reiterate, in non-babbling equilibria, committee members do not ignore the information contained in the messages. There always exist babbling equilibria with truthful communication in Majority, namely equilibria where subjects always vote R regardless of messages, where unilateral attempts to respond to the messages are payoff irrelevant, see Lemma 4.

experiment, 10/40 and 15/35, are comparably small in the sense that they do not dominate the social payoffs from making the socially optimal decision, but are large enough to suppress truthful communication in Majority. That is, both Majority treatments satisfy $K \in \{\hat{\Gamma}\}^K$ as referred to in Corollary 1, which yields the following research question.

Question 1. *Are messages more likely to be truthful under Consensus than under Majority, and does it depend on the private signal being R or B?*

We predict *yes* to both parts. More specifically, Proposition 1 shows that Majority may induce asymmetric misreporting since agents best-respond to truthful reporting by sending truthful messages given a signal of *B*, and misreport given a signal of *R* to induce their co-players to vote for *B*. Additionally, Lemma 2 predicts that agents who receive a signal of *R* and misreport will vote *R*, given that their co-players vote for *B* with a higher probability. We refer to this as the free-rider problem.

As a result of this free-rider problem, Corollary 1 predicts that Consensus will outperform Majority in terms of information aggregation. Additionally, we predict that information aggregation will disproportionately fail when the aggregate profile of signals indicates that the committee should select *B*: If a majority of subjects receive *B* signals, then the free-rider problem implies that *R* will be selected with positive probability. If a majority of subjects receive *R* signals, however, then the committee will select *R* with high probability since subjects playing a free-rider strategy will vote *R* regardless of the messages. Therefore, our model suggests the following research question.

Question 2. *Is information aggregated more efficiently under Consensus, and does it depend on the committee-optimal decision being R or B?*

Conditional on a positive finding for Questions 1 and 2, theoretically, the channel for the increased level of information aggregation under Consensus could be twofold. On the one hand, committee members may anticipate and account for the more truthful messages under Consensus. On the other hand, even if subjects hold the belief that their co-players' messages are truthful, Lemma 2 shows that Majority induces a coordination problem in the voting stage that may still prevent optimal information aggregation.

Therefore, we continue our analysis by identifying the mechanisms behind any relative decrease in information aggregation under Majority. First, we consider the question of whether subjects account for the average level of truthful reporting in their voting decisions.

Question 3. *Do subjects anticipate and respond to more truthful messages under Consensus?*

Second, Lemma 2 and Proposition 1 predict that agents will best-respond to truthful reporting by free-riding, misreporting *B* to free ride on the *B* votes of their co-players, while personally voting for R. This prediction provides us with an opportunity to verify the causal mechanism predicted by our model.

Question 4. *Do subjects strategically misreport under Majority to free-ride?*

If the findings of the experiment negate this prediction, then subjects may be falsely reporting their signals for reasons other than what we predict, and hence our model would be falsified.

5 Analysis of the experiment

We address the four research questions successively. Table 2 provides a first overview of the experimental results, delineating the truthfulness of messages and the level of information aggregation by decision rule, expressive payoffs and signal.

Table 2: Average treatment effects in relation to Questions 1 and 2

Expressive payoff	Low		High	
Decision Rule	Majority	Consensus	Majority	Consensus
Truthful message if R signal	90%	96%***	79%	94%***
Truthful message if B signal	84%	86%	87%	86%
Committee-optimal decision R	80%	79%	88%	86%
Committee-optimal decision B	61%	83%***	60%	83%***

Mann-Whitney-Wilcoxon test (Consensus = Majority); * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

5.1 Are messages more truthful under Consensus?

The upper two lines of Table 2, and similarly Figure 1, provide information on the truthfulness of communication, disaggregated by treatment and signal as required to answer Question 1. The asymmetric prediction is satisfied for both signal R and signal B: Given a signal of R, in both the Low and High expressive payoff conditions, messages are significantly more truthful under Consensus than under Majority. That is, with a signal of R, misreporting increases from 4% to 10% given Low expressive payoffs, and from 6% to 21% given High expressive payoffs.

In contrast, truthful reporting given a signal of B is very stable across all conditions: The differences between Consensus and Majority are very small (at most two percentage points) and far from being significant, confirming the second part of Question 1. Across all treatments, the average rate of truthful reporting, roughly 85 percent, is lower than might be expected given the results of Guarnaschelli et al. (2000) (their experiment considers homogeneous payoffs,

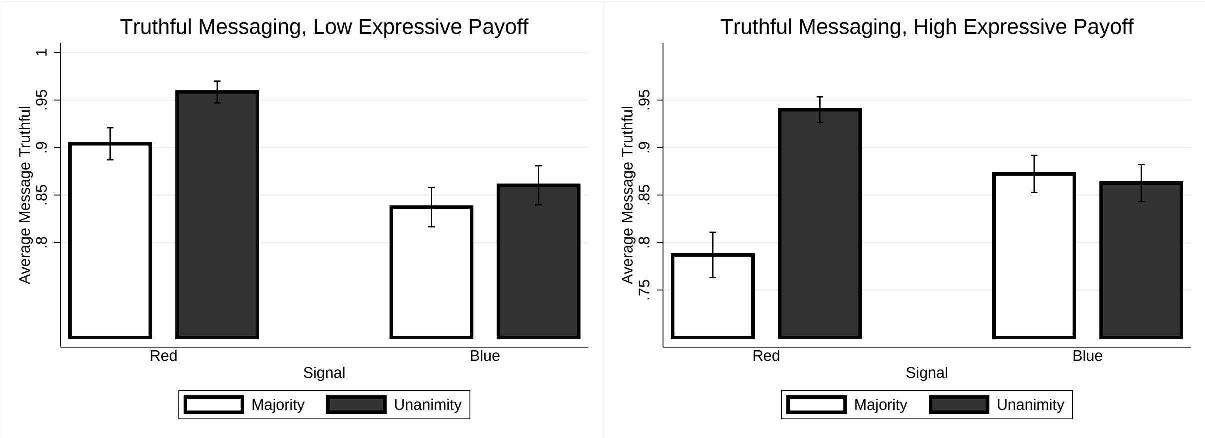


Figure 1: Average levels of truthful messaging ($m_i = s_i$) by signal and treatment.

but is otherwise comparable). Focusing only on Consensus, however, the relative frequency of truthful messages is comparable with Guarnaschelli et al. (2000).

Result 1. *As predicted (Question 1), messages are more truthful in Consensus after R signals and equally truthful after B signals.*

A related point apparent in Figure 1 is worthy of mention. Under Majority, the level of misreporting given a signal of R more than doubles for the High vote-contingent payoff treatment, from 10% to 21%. This finding is consistent with the quantitative increase in the incentive to strategically misreport given a signal of R. In contrast, the average level of truthful reporting by signal is stable across the Low and High treatments under Consensus. This finding corroborates Result 1 and suggests that, under Majority, the level of truthful communication is sensitive to the size of the incentive to engage in strategic misreporting. Section B in the Appendix further shows that the results are highly robust to experience.

5.2 Is information aggregated more efficiently under Consensus?

While truthful communication is a necessary condition for the committee to behave optimally, the most pertinent comparison between Majority and Consensus is the ability of the committee to efficiently aggregate the private information of its members. Figure 2 shows the committee decision as a function of the aggregate profile of signals, and the lower two lines in Table 2 provide the relative frequencies of committee-optimal decisions, disaggregated by whether R or B is optimal — across all conditions examined here, the committee-optimal decision is B if and only if all subjects have received B signals. As detailed in Question 2, our prediction is that there is no difference between Consensus and Majority if the expressive option, R, is committee-optimal, but that under Consensus the committee will be more likely to select the non-expressive option, B, when it is optimal.

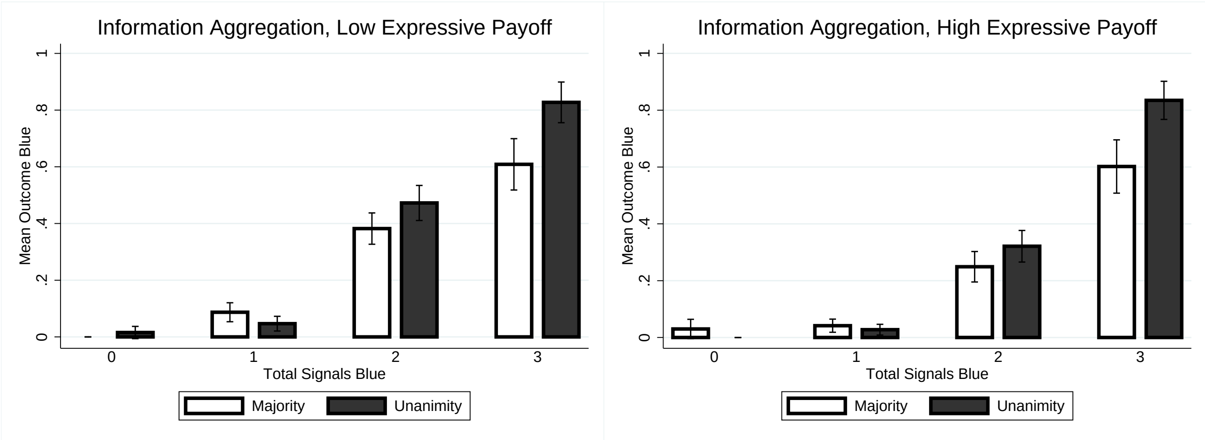


Figure 2: Average levels of outcome *B* as a function of the aggregate profile of *signals* (not messages) by treatment.

The experimental results are very sharp. If *R* is committee-optimal, there is virtually no difference in the probability that the committee selects *R* between Consensus and Majority: 80% versus 79% for Low, and 88% versus 86% for High expressive payoffs. If *B* is committee-optimal, the difference is large and highly significant: 22 percentage points (61% versus 83%) for Low expressive payoffs, and 23 percentage points (60% versus 83%) for High expressive payoffs. In all, the probability that the committee correctly aggregates information given three signals for *B* increases by over a third under Consensus.[11]

Result 2. *As predicted (Question 2), the committee is significantly more likely to select the optimal option under Consensus if the non-expressive option, B, is committee-optimal, and the difference between the two decision rules is negligible if R is optimal.*

5.3 Do subjects anticipate more truthful messages under Consensus?

Having established that Consensus results in more truthful messaging and increased information aggregation (Questions 1 and 2) we now turn to the more subtle questions regarding mechanism (Questions 3 and 4). For a first pass at exploring the effect of messages under the different decision rules, we estimate a discrete choice model that considers each individual's voting decision as a function of the voting rule, and the information known by the subjects at the time they take their voting decision. Table 3 summarizes the estimation.

The regression results indicate that subjects respond to both an own signal of *B* and to co-player's messages for *B* by voting for *B* with a higher probability. Moreover, under Consensus,

[11] The difference between the two decision rules is also slightly larger in the 2nd half of the experiment (see Section B in the supplementary appendix). Additionally, note that subjects are more likely to select *B* given two signals of *B* under Consensus, although this difference is attenuated in the second half of the experiment.

the impact of a message for B from a co-player has a similar the impact on voting behavior as an own signal for B. Under Majority, however, subjects are significantly less likely to respond to co-players' messages of B, relative to their own signal (see the negative coefficient on the interaction of "Majority" and "Other's messages"). This strongly suggests that the higher proportion of misreporting observed under Majority negatively impacts the committee's ability to efficiently aggregate the information of its members.

Table 3: Probit estimations to explain individual votes/opinions for B

Vote/Opinion B	(1) Low	(2) High	(3) Joint
Own signal B	0.299*** (0.0246)	0.282*** (0.0174)	0.291*** (0.0148)
Number of other's messages B	0.361*** (0.0128)	0.298*** (0.0112)	0.328*** (0.00910)
Majority	0.0884** (0.0262)	0.0624* (0.0252)	0.0732*** (0.0185)
Majority × own signal	−0.0121 (0.0373)	0.00908 (0.0361)	−0.00168 (0.0256)
Majority × other's messages	−0.138*** (0.0255)	−0.109*** (0.0218)	−0.121*** (0.0171)
High			−0.0689*** (0.0171)
Constant	−0.163*** (0.0109)	−0.171*** (0.00893)	−0.132*** (0.0107)
N	4650	4650	9300

Subject-level clustered standard errors in parentheses; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

This direct comparison between the decision rules is suggestive of the predicted result. However, we can make the analysis more precise by controlling for differences in expected payoffs from voting B under the two decision rules. That is, while we see an increased responsiveness to co-players' messages under Consensus, the choice environment in the voting stage is not directly comparable to Majority. Therefore, we supplement the finding of the discrete choice model by investigating the impact of co-player's messages on the subjects' implied beliefs, taking into account the actual differences in expected payoffs.

In line with Question 3, our hypothesis is that subjects' beliefs are more responsive to co-players' messages under Consensus. We test this hypothesis by estimating a simple structural equation model of the decision making process.[12] The equation system directly implements the strategic game played by the subjects. First, a subject's belief about the true state $\omega \in \{R, B\}$ is

[12] An arguably simpler approach would be to plainly ask subjects about their beliefs, but in the context of beliefs underlying strategic decisions, the elicited beliefs have been found to be incompatible with the chosen actions (Costa-Gomes and Weizsäcker, 2008). Even without such obstacles, robustly incentive-compatible elicitation of beliefs, prior and after revelation of messages, is not simple either and may distract or appear obtrusive to subjects.

a function of the own signal $s \in \{R,B\}$ and the opponents' messages $m_2, m_3 \in \{R,B\}$,

$$\Pr(\omega = R | s, m_2, m_3) = \frac{1}{1 + \exp\{\alpha_1 (I_{s=B} - 0.5) + \alpha_2 (I_{m_2=B} + I_{m_3=B} - 1)\}} \tag{2}$$

with belief parameters α_1, α_2. Here we use the indicators $I_{s=B}, I_{m_2=B}, I_{m_3=B}$ to indicate whether the own signal (s) or the opponents' messages (m_2, m_3) are equal to B (value 1) or not (value 0).

Second, a subject's belief about the voting outcome $X \in \{R,B\}$, conditional on the own vote $v \in \{R,B\}$ and the number of B messages, is[13]

$$\Pr(X = R | v_i) = \frac{1}{1 + \exp\{\beta_1 \cdot I_{v=B} + \beta_{2|m_1+m_2+m_3}\}} = 1 - \Pr(X = B | v_i). \tag{3}$$

Here, β_1 captures the weight of the own vote, and $\beta_{2|\cdot}$ captures the expectation about the opponents' votes as a function of the message profile. Specifically, $\beta_{2|\cdot}$ is a vector of four values, where $\beta_{2|0}$ denotes the expectation in the case where there are zero B messages, and $\beta_{2|1}$, $\beta_{2|2}$, $\beta_{2|3}$ concern the cases of one, two, or three B messages. In conjunction, these two beliefs define the probability that the voting outcome X is equal to the true state ω,

$$\Pr(X = \omega | v_i) = \Pr(\omega = R) \cdot \Pr(X = R \,|\, v) + \Pr(\omega = B) \cdot \Pr(X = B \,|\, v).$$

Finally, using P_c and K to denote the payoff from the voting outcome being correct ($X = \omega$) and the expressive payoff from voting R, respectively, voting for R has probability

$$\Pr(v = R) = \frac{1}{1 + \exp\{-\lambda \cdot P_c \cdot (\Pr(X = \omega | v_i = R) - \Pr(X = \omega | v_i = B)) - \lambda \cdot K\}},$$

allowing for logistic errors (with precision parameter $\lambda \geq 0$). Note that this model is fairly general. Depending on how the belief parameters $(\alpha_1, \alpha_2, \beta_1, \beta_{2|\cdot})$ relate to their empirical counterparts, the model is compatible with (ir)rational expectations, overshooting in Bayesian updating, cursed beliefs and level-k beliefs. The empirical counterparts $(\hat{\alpha}_1, \hat{\alpha}_2, \hat{\beta}_1, \hat{\beta}_{2|\cdot})$ can be estimated simply by logit regressions. The rational signal and message weights $\hat{\alpha}_1$ and $\hat{\alpha}_2$ are estimated by regressing the true state of the world ω on the signal and messages, using Eq. (2), and the rational outcome weights $\hat{\beta}_1$ and $\hat{\beta}_{2|\cdot}$ are estimated by regressing the probability that the correct decision is made ($X = \omega$) on the own vote and the message profile, using Eq. (3).

We control for expected payoffs as observed in the experiment, while still allowing for two forms of behavioral biases, by requiring (1) the belief weights α_1, α_2 to be jointly proportional to their rational expectation counterparts, which allows for overshooting (base rate fallacy) or conservative belief formation as observed in many experiments, and (2) the vote weights $\beta_{2|\cdot}$ to be jointly proportional to their empirical counterparts, which allows for over- or underestimat-

[13] Slightly abusing notation, we use $m_1 + m_2 + m_3$ as shortcut for $I_{m_1=B} + I_{m_2=B} + I_{m_3=B}$.

ing the predictability of the co-players' votes. In the extreme case, $\beta_{2|\cdot} = 0$, subjects believe their co-players are perfectly unpredictable (level-1, Stahl and Wilson, 1995). If $0 < \beta_{2|\cdot} < \hat{\beta}_{2|\cdot}$, subjects underestimate the predictability of others as observed by Weizsäcker (2003), Goeree and Holt (2004) and Eyster and Rabin (2005), and if $\beta_{2|\cdot} = \hat{\beta}_{2|\cdot}$, subjects hold rational expectations of the mapping from their co-players' messages to votes. We report on two robustness checks invoking rational expectations in either dimension below, but the results are very robust in general.

We find that subjects overshoot in belief formation, given signals and messages, and therefore hold rather strong beliefs. To see this effect as clearly as possible, it is best to look at cases where the co-players' messages contradict the own signal. That is, we look at beliefs about the true state ω after a private B signal and two R messages of co-players, and after a private R signal and two B messages of co-players. Given the above notation, we hypothesize that beliefs react more strongly to the co-players' messages under Consensus, since messages are more truthful in this case; i.e.,

$$\begin{aligned}
\Pr_{\text{Maj}}(\omega = R|s = B, m_2 = m_3 = R) &< \Pr_{\text{Una}}(\omega = R|s = B, m_2 = m_3 = R), \\
\Pr_{\text{Maj}}(\omega = R|s = R, m_2 = m_3 = B) &> \Pr_{\text{Una}}(\omega = R|s = R, m_2 = m_3 = B).
\end{aligned}$$

Based on the estimates of the structural equation models, namely α_1, α_2 in conjunction with Eq. (2), these beliefs can be computed straightforwardly. Since the beliefs are based on estimates of α_1, α_2, we bootstrap their distributions to test our hypotheses (resampling at the subject level to account for the panel character of the data, stratifying to acknowledge the treatment structure). Parameters are estimated by maximum likelihood[14] and both the standard errors of the parameters as well as the p-values of the null hypotheses are also bootstrapped.

The results are reported in Table 4. First, looking at the empirically true probabilities, we can see that in all cases, subjects should tend to follow the opponents' messages when both are the same despite contradicting the own signal. For example, in the second halves of the sessions, after an R signal and two B messages from the opponents, the empirical probabilities that the state is R are 43.5% and 37% under Majority and Consensus, respectively. This shows that messages should be given weight—and more so under Consensus treatments than under Majority, as the empirical probabilities deviate relatively more from 50-50 under Consensus. The remaining columns of Table 4 show that, in all cases, subjects' beliefs indeed deviate more from 50-50, in the direction of the messages, in consensus treatments than in majority treatments. The differences are significant, obtain robustly in both the first and the second halves of the sessions, and the robustness checks assuming rational expectations forming either state or outcome beliefs confirm the result. Based on this, we conclude that subjects anticipate

[14] To maximize the likelihood, we first use the gradient-free NEWUOA approach (Powell, 2006), which is comparably robust (Rios and Sahinidis, 2013), and secondly a Newton-Raphson algorithm to ensure convergence.

Table 4: Structural equation analysis of beliefs ($\omega = Red$) as function of signals and messages

	Empirical	Baseline	Rational 1	Rational 2
First halves of sessions				
Beliefs after A signal and two B messages from others				
$\Pr_{\text{Maj}}(\omega = R \mid s = R, m_2 = m_3 = B)$	0.476 (0.033)	0.407 (0.011)	0.478 (0)	0.394 (0.012)
$\Pr_{\text{Una}}(\omega = R \mid s = R, m_2 = m_3 = B)$	0.452 (0.031)	0.328^{--} (0.018)	0.461^{--} (0)	0.361^{-} (0.011)
Beliefs after B signal and zero B messages from others				
$\Pr_{\text{Maj}}(\omega = R \mid s = B, m_2 = m_3 = R)$	0.505 (0.03)	0.593 (0.011)	0.522 (0)	0.606 (0.012)
$\Pr_{\text{Una}}(\omega = R \mid s = B, m_2 = m_3 = R)$	0.559 (0.029)	0.672^{++} (0.018)	0.539^{++} (0)	0.639^{+} (0.011)
Log-Likelihood		−2021.5	−2280.2	−2312.4
Robustness checks				
Rational state beliefs			✓	
Rational voting beliefs				✓
Second halves of sessions				
Beliefs after A signal and two B messages from others				
$\Pr_{\text{Maj}}(\omega = R \mid s = R, m_2 = m_3 = B)$	0.435 (0.029)	0.397 (0.012)	0.465 (0)	0.378 (0.013)
$\Pr_{\text{Una}}(\omega = R \mid s = R, m_2 = m_3 = B)$	0.37 (0.035)	0.321^{--} (0.016)	0.422^{--} (0)	0.336^{--} (0.009)
Beliefs after B signal and zero B messages from others				
$\Pr_{\text{Maj}}(\omega = R \mid s = B, m_2 = m_3 = R)$	0.527 (0.031)	0.603 (0.012)	0.535 (0)	0.622 (0.013)
$\Pr_{\text{Una}}(\omega = R \mid s = B, m_2 = m_3 = R)$	0.551 (0.027)	0.679^{++} (0.016)	0.578^{++} (0)	0.664^{++} (0.009)
Log-Likelihood		−1850.2	−2019.6	−2075.5
Robustness checks				
Rational state beliefs			✓	
Rational voting beliefs				✓

Note: In the baseline model we allow for mistakes in Bayesian updating when forming state beliefs and voting beliefs. That is, the respective belief parameters (α_1, α_2) and $\beta_{2|\cdot}$ are allowed to be arbitrarily scaled vectors of their rational expectation counterparts $(\hat{\alpha}_1, \hat{\alpha}_2)$ and $\hat{\beta}_{2|\cdot}$ as estimated from logistic regressions. The robustness checks enforce rational expectations by equating the respective belief parameters with their rational expectation counterparts. The respective likelihoods are significantly worse than that of the baseline model (showing that subjects do actually not hold rational expectations), but the main result that implicit beliefs differ between majority and consensus treatments is robust nonetheless. As before, significance of differences between majority and consensus estimates is indicated by plus and minus signs ($^{++/--}$ at $p < .05$ and $^{+/-}$ at $p < .1$ using bootstrapped p-values), next to the consensus treatment estimates. All standard errors are bootstrapped.

and account for the higher probability of truthful messages under Consensus.

Result 3. *Subjects' beliefs react more strongly to co-players' messages under Consensus, showing that they respond to the increased truthfulness of messages.*

5.4 Do subjects strategically misreport to free-ride?

Lastly, we explore the theoretical prediction that subjects will best-respond to truth-telling by misreporting and free-riding. For suggestive evidence regarding free-riding, it is instructive to consider the average voting strategy of subjects the case of three messages for B ($M^{\#} = 3$). In this case, for the Majority/High expressive payoff treatment where misreporting is the most common, subjects who misreport their signal vote for B just 16 percent of the time, relative to 69 percent for subjects who sent a truthful message of B (this is also much lower to the analogous rate under Consensus/High, which is 54 percent; see Table 5 in the appendix). This low rate of voting for B is consistent with the free-riding strategy of misreporting B and then voting for R.

To identify whether Majority causes subjects to play the "free-riding" strategy, however, we need to link messages and voting at the individual level. Moreover, free-riding is not the only reason to misreport signals. Subjects may hold different beliefs as to which option maximizes expected payoffs conditional on a given set of signals. Accordingly, a subject with a bias for B, for example, may choose to misreport their signal to increase the probability that the committee chooses B given two signals for B (*persuasive misreporting*).

Overall, it is easy to think of at least five classes of individual strategies, summarized in Table 5a. In addition to the three strategic types outlined above, "strategic Red/Blue" as persuasive misreporting and "free-riding" as strategic misreporting, we also allow for subjects who are honest in the messaging stage and believe that other agents message honestly ("honest/naive"), and subjects with noisy behavior as a residual family to collect the players that do not fit into either of the other four classes ("noisy"). Such apparently noisy behavior may result from either misunderstanding the game or, more likely, from playing inconsistently over the course of the session. The inclusion of the noisy type thus serves as a robustness check of both the statistical classification and of the adequacy of the experimental set-up more generally. Our objective will be to evaluate our prediction that subjects more frequently use honest/naive strategies in consensus treatments and free-riding strategies in majority treatments.

The voting strategies we assign to the honest/naive type follow from the theoretical predictions of Lemma 2. Specifically, the honest type will vote for R given two or more messages/signals for R. With two messages/signals for B, the honest type will vote for B with an intermediate probability, and with three messages/signals for B they will vote for B with a high probability. The messaging and voting strategies of the strategic types are then assigned relative to the honest type: The strategic Red type is more likely to message and vote R, while the strategic Blue type is more likely to message and vote B. The free-rider type, on the other hand, is more likely to message B and vote R.[15] The detailed definitions are provided in Table 5a. We allow for $\{\pi_{Lie}, \pi_{Low}, \pi_{Medium}, \pi_{High}\}$ to be free parameters in the estimation to avoid an inadequate specification.

A statistically efficient approach to determine the latent weights of the strategy classes is finite mixture modeling (MacLachlan and Peel, 2000). Such mixture models have been commonly used in behavioral analyses since Stahl and Wilson (1994, 1995), and have more recently been used, for example, to study individual heterogeneity in dictator games (Cappelen et al., 2007) and individual behavior in repeated games (Dal Bó and Fréchette, 2011; Fudenberg et al., 2012; Breitmoser, 2015). Finite mixture modeling is a general approach allowing for probabilistic assignment of subjects to strategy classes, which resolves a number of concerns with deterministic assignments,[16] but is otherwise comparable to cluster analyses.[17]

We assume that ex-ante, a subject plays strategy $k \in K$ with probability ρ_k, and that each subject sticks with the chosen strategy throughout the analyzed interactions. The statistical model is fully described by the ex-ante strategy weights $\rho = (\rho_1, \rho_2, \ldots)$ and the strategy parameters $\pi = (\pi_{\text{Lie}}, \pi_{\text{Low}}, \pi_{\text{Med}}, \pi_{\text{High}})$ discussed above. Formally, given a subject pool $S^{\#}$ and our set of observations $O = \{o_s\}_{s \in S}$, let $P(o_s|\pi, k)$ denote the probability of the choices o_s made by subject $s \in S$ assuming s plays strategy of type k with parameters π. Then, the likelihood function

$$LL(\rho, \pi \,|\, O) = \sum_{s \in S} \log \sum_{k \in K} \rho_k \cdot P(o_s|\pi, k)$$

is maximized over (ρ, π) to estimate the ex-ante strategy weights ρ we are interested in. The strategy parameters π are not of direct interest for our research hypothesis, but allow us to test whether the estimates align with our ex-ante predictions, which also serves as a robustness check. Given the observed choices O, we can determine the posterior class assigment of each subject $s \in S$ simply by applying Bayes Rule.

This approach does not require us to commit to distance functions and expresses the degree of (un)certainty as a function of behavior by implying probabilistic posterior beliefs. As usual, we maximize the likelihood by the expectation-maximization (EM) algorithm (see e.g.

[15] While the relative comparisons follow from the theory, the exact division into the strategy classes was calibrated using the aggregate voting strategies reported in Table 6 in the Appendix.

[16] Deterministic approaches that assign each subject to a strategy class based on some distance measure are sensitive to the distant measure chosen, are ambivalent near the boundaries of each class, and do not reflect the degree of (un)certainty about a subject's classification. Furthermore, deterministic classification requires the distances to be reliably measured. In our case, however, they would be based on only few observations per information set.

[17] In most behavioral cluster analyses, each data point (subject) is represented by vectors with few elements. In such cases, we can plot the individual estimates and "mark" the cluster areas. This approach is inadequate here, as each subject is characterized by choices in many different information sets (namely, fourteen) with comparably few observations in each case. Loosely speaking, the finite mixture analysis determines a common denominator across information sets with regards to strategies.

Table 5: Do subjects play more honest strategies in consensus voting? Results of the mixture analysis

(a) Definition of the classes of strategies

	Messages		Voting											
	$\mu(A)$	$\mu(B)$	$\pi(A,A,0)$	$\pi(B,A,0)$	$\pi(A,A,1)$	$\pi(A,B,1)$	$\pi(B,B,1)$	$\pi(B,A,1)$	$\pi(A,A,2)$	$\pi(A,B,2)$	$\pi(B,B,2)$	$\pi(B,A,2)$	$\pi(A,B,3)$	$\pi(B,B,3)$
Noise	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5	.5
Honest	1	0	1	1	1	1	1	1	π_{Med}	1	π_{Med}	π_{Low}	π_{Med}	π_{Low}
StratRed	1	π_{Lie}	1	1	1	1	1	1	1	1	1	π_{Med}	1	π_{Med}
StratBlue	$1-\pi_{\text{Lie}}$	0	1	1	1	1	1	π_{Med}	π_{Med}	1	π_{Med}	π_{Low}	π_{Med}	π_{Low}
Freeride	$1-\pi_{\text{Lie}}$	0	1	1	1	1	1	π_{High}	π_{High}	1	π_{High}	π_{Med}	π_{High}	π_{Med}

Note: $\mu(s)$ is the probability of sending message A given the signal $s \in \{A,B\}$. $\pi(s,m,M)$ is the probability of voting A as a function of one's signal s, message m, and the number $M^{\#}$ of B messages overall (i.e. in aggregate over all players). The parameters $(\pi_{Lie}, \pi_{\text{Low}}, \pi_{\text{Med}}, \pi_{\text{High}})$ allow adaptation to subjects' behavior, with the theoretical ex-ante hypothesis $\pi_{\text{Low}} < \pi_{\text{Med}} < \pi_{\text{High}}$.

(b) Strategy weights and parameters across treatments (bootstrapped standard errors in parentheses)

	Strategy weights in population					Strategy parameters					
	Noise	Honest	StratRed	StratBlue	FreeRide	ε	π_{Lie}	π_{High}	π_{Med}	π_{Low}	ICL-BIC
All games per session											
Majority 35 15	0.12	0.44	0.19	0	0.26	0.04	0.42	0.73	0.35	0.08	6368.79
	(0.05)	(0.11)	(0.07)	(0.01)	(0.09)	(0)	(0.11)	(0.06)	(0.04)	(0.03)	
Majority 40 10	0.07	0.45	0.25	0.03	0.2						
	(0.04)	(0.12)	(0.08)	(0.03)	(0.07)						
Unanimity 35 15	0.07	0.78^{++}	0.11	0.04	0^{--}						
	(0.04)	(0.06)	(0.05)	(0.03)	(0)						
Unanimity 40 10	0.04	0.75^{++}	0.18	0.02	0^{--}						
	(0.03)	(0.07)	(0.06)	(0.03)	(0.02)						
Majority	0.09	0.44	0.22	0.01	0.23	0.04	0.41	0.72	0.35	0.08	6345.24
	(0.03)	(0.1)	(0.06)	(0.02)	(0.06)	(0)	(0.11)	(0.05)	(0.04)	(0.03)	
Unanimity	0.06	0.77^{++}	0.15	0.03	0^{--}						
	(0.03)	(0.05)	(0.04)	(0.02)	(0.01)						
Robustness check 1: 1st halves per session											
Majority	0.13	0.59	0.12	0.05	0.12	0.04	0.61	0.79	0.36	0.14	3304.43
	(0.04)	(0.11)	(0.06)	(0.04)	(0.08)	(0)	(0.13)	(0.13)	(0.04)	(0.05)	
Unanimity	0.1	0.78^{++}	0.12	0	0^{--}						
	(0.03)	(0.04)	(0.03)	(0.01)	(0)						
Robustness check 2: 2nd halves per session											
Majority	0.1	0.48	0.19	0.01	0.22	0.03	0.41	0.86	0.38	0.06	3009.04
	(0.03)	(0.15)	(0.06)	(0.02)	(0.11)	(0.01)	(0.2)	(0.06)	(0.09)	(0.05)	
Unanimity	0.07	0.71^{++}	0.17	0.04	0^{--}						
	(0.03)	(0.08)	(0.06)	(0.02)	(0.02)						

Note: This table provides the statistical support for our observation that subjects use more honest/naive strategies in consensus treatments and more freeriding strategies in majority treatments. The table reports the weights of the five predicted strategies in the population, the estimated strategy parameters $(\pi_{Lie}, \pi_{\text{Low}}, \pi_{\text{Med}}, \pi_{\text{High}})$, the bootstrapped standard errors, and the goodness-of-fit measures ICL-BIC. The upper panel provides the estimates for the entire sessions, the lower panel provides robustness checks focusing on either first halves and second halves of the sessions. In the upper panel, we report estimates distinguishing either all treatments or only majority and consensus treatments. Plus and minus signs indicate significant differences ($^{++}$ at $p < .05$ and $^{+}$ at $p < .1$ using bootstrapped p-values) of the strategy weights in the consensus treatments compared to weights in the respective majority treatments. The ICL-BICs show that the latter more parsimonious approach is statistically more adequate, but the main results are robust in either case. They also hold robustly if we focus on either the first or the second halves of the sessions, as shown in the lower panel.

Arcidiacono and Jones, 2003), and in the maximization step we again first use the gradient-free NEWUOA approach and secondly a Newton-Raphson algorithm to ensure convergence. Model adequacy is measured using ICL-BIC (Biernacki et al., 2000), which penalizes both superfluous model components and excessive parametrization. ICL-BIC has been shown to enable reliable estimation of the number of components (in our case, strategy classes) in the population (Fonseca and Cardoso, 2007). Finally, standard errors are bootstrapped by replacement at the subject level to account for the panel character of the data, using stratified resampling acknowledging the treatment structure.

Table 5b presents the estimated strategy weights and strategy parameters, alongside the bootstrapped standard errors and statistical tests of our hypothesis. The main results are that 44 and 45% of the subjects use honest/naive strategies in the Majority treatments, compared to 78 and 75% in the Consensus treatments, and that 26 and 20% of the subjects free-ride in the Majority treatments but no subjects free-ride in the Consensus treatments.

The respective differences between Majority and Consensus treatments are all highly significant and as hypothesized. The result is robust to pooling the Majority treatments and Consensus treatments, respectively, and robust to focusing on either the first halves or second halves of each session, as shown in the lower panel of Table 5b. Further, the strategic parameters satisfy $\pi_{\text{Low}} < \pi_{\text{Med}} < \pi_{\text{High}}$, showing that the subjects use the strategies as predicted, and the share of unclassified ("noisy") players is around or below 10%, showing that subjects use their respective strategies consistently throughout the session. Finally, in our robustness checks reported in Table 8 (see Appendix B), we find that none of the strategy classes are superfluous, although some weights are small, in the sense that eliminating either class increases the ICL-BIC measure of model adequacy. With these robustness checks in mind, we conclude as follows.

Result 4. *Subjects use honest/naive strategies more frequently in Consensus and strategically misreport to free-ride in Majority.*

5.5 Discussion

Our experimental results show that expressive payoffs lead to strategic communication and inefficient information aggregation when committees take decisions via Majority rule. Additionally, relative to Consensus, we demonstrate that subjects are less responsive to other subjects' messages under Majority, both in terms of the voting decision and their implicit ex-post beliefs regarding the state of the world. We find evidence that this decrease in the effectiveness of communication is due to the fact that, under Majority, a subset of subjects adopt a "free-riding" strategy, falsely reporting the non-expressive option to encourage other subject to vote for this option, while personally voting for the expressive option.

It is important to note, however, that both a coordination and a free-rider problem exist under Majority even with truthful communication—the coordination problem may contribute to the decrease in information aggregation relative to Consensus. This raises the question of how our results might change in an environment with a richer communication set. For example, Goeree and Yariv (2011) show that given access to free-form communication, the most prevalent (endogenous) procedure by which subjects take decisions is to (1) share their private signals, (2) coordinate on a committee outcome, and (3) unanimously vote for the outcome decided on in (2). In contrast, we constrain subjects to a binary message, effectively ruling out the coordination stage (2).

Generally, communication has been shown to be an effective tool for coordination (see Ledyard, 1995 for an overview), and one might expect that a richer message space may improve the committee's ability to select the committee-optimal option. However, in contrast to the setting analyzed in Goeree and Yariv (2011), a unanimous vote for the non-expressive option is not an equilibrium with expressive payoffs, since subjects can deviate to the expressive option while leaving the committee decision unchanged. Therefore, the impact of a richer communication space on *information aggregation* is ambiguous in our setting: On one hand, a richer communication space may aid in coordinating on choosing the non-expressive option when the profile of messages indicates that this option is optimal. On the other hand, greater coordination increases the incentive to behave strategically: if the probability that other subjects vote for B given three messages of B increases, then the incentive to deviate from truthful communication to the free-riding strategy increases (moreover, our experiment shows that the severity of free-riding increases with incentives). That is, with increased coordination, a subject with signal R who deviates from an honest strategy to free-ride (message B and vote R) faces a higher probability that the committee chooses B given only two signals for B.[18] These conflicting mechanisms highlight richer communication in a setting with expressive payoffs as a promising area for further research.

6 Conclusion

In this paper, we generalize the standard model of information aggregation through voting to account for the possibility that committee members receive vote-contingent payoffs. Using a theoretical model, we show that when the committee aggregates votes via a majority rule, truthful communication and informative voting is not an equilibrium despite the fact that committee members have homogeneous preferences over the committee outcome. In contrast, an equilibrium with truthful communication and committee-optimal voting strategies is an equi-

[18]Note that a similar concern emerges when considering an increase in the committee's size: A larger committee gives access to more signals, potentially increasing the accuracy of the committee's decision. However, given voting B is a public good, consistent with the evidence on free-riding and population size (see Isaac and Walker, 1988), a larger committee may induce more subjects to adopt a free-riding strategy.

librium under a two-sided unanimity (consensus) rule as long as committee members have homogeneous preferences over the committee outcome conditional upon the aggregate profile of signals. This finding contrasts with previous results in the literature, which have suggested that unanimity is a uniquely bad decision rule for aggregating information, and suggests a novel rationale for the use of a consensus rule: in settings with vote-contingent payoffs, efficiency can only be assured under a decision rule that uniformly enforces responsibility for the committee decision across all committee members.

We test the predictions of the model using laboratory experiments. Our experimental results broadly support for the theoretical predictions. We find that, relative to a consensus rule, subjects are more likely to falsely report their signal and committee decisions are less likely to aggregate private information under majority rule. Moreover, we identify that this decrease in information aggregation can be attributed to subjects adopting a "free-rider" strategy under majority, which leads to less effective communication and sub-optimal committee decisions.

References

Arcidiacono, P. and Jones, J. B. (2003). Finite mixture distributions, sequential likelihood and the em algorithm. *Econometrica*, 71(3):933–946.

Austen-Smith, D. and Banks, J. (1996). Information Aggregation, Rationality, and the Condorcet Jury Theorem. *American Political Science Review*, 90(1):34–45.

Austen-Smith, D. and Feddersen, T. J. (2006). Deliberation, preference uncertainty, and voting rules. *American Political Science Review*, 100(02):209–217.

Biernacki, C., Celeux, G., and Govaert, G. (2000). Assessing a mixture model for clustering with the integrated completed likelihood. *IEEE transactions on pattern analysis and machine intelligence*, 22(7):719–725.

Bouton, L., Llorente-Saguer, A., and Malherbe, F. (2017a). Get Rid of Unanimity Rule: The Superiority of Majority Rules with Veto Power. *Journal of Political Economy, Forthcoming*, Forthcoming.

Bouton, L., Llorente-Saguer, A., and Malherbe, F. (2017b). Unanimous rules in the laboratory. *Games and Economic Behavior*, 102:179–198.

Breitmoser, Y. (2015). Cooperation, but no reciprocity: Individual strategies in the repeated prisoner's dilemma. *The American Economic Review*, 105(9):2882–2910.

Brennan, G. and Buchanan, J. (1984). Voter choice: Evaluating political alternatives. *American Behavioral Scientist*, 28(2):185–201.

Canes-Wrone, B., Brady, D., and Cogan, J. (2002). Out of step, out of office: Electoral accountability and House members' voting. *American Political Science Review*, 96(1):127–140.

Cappelen, A. W., Hole, A. D., Sørensen, E. Ø., and Tungodden, B. (2007). The pluralism of fairness ideals: An experimental approach. *The American Economic Review*, 97(3):818–827.

Chan, J., Lizzeri, A., Suen, W., and Yariv, L. (2017). Deliberating Collective Decisions. *Review of Economic Studies*, Forthcoming.

Costa-Gomes, M. A. and Weizsäcker, G. (2008). Stated beliefs and play in normal-form games. *The Review of Economic Studies*, 75(3):729–762.

Coughlan, P. (2000). In Defense of Unanimous Jury Verdicts: Mistrials, Communication, and Strategic Voting. *American Political Science Review*, 94(2):375–393.

Dal Bó, P. and Fréchette, G. R. (2011). The evolution of cooperation in infinitely repeated games: Experimental evidence. *The American Economic Review*, 101(1):411–429.

de Condorcet, M. (1785). *Essai sur lapplication de lanalyse à la probabilité des décisions rendues à la pluralité des voix'*. L'imprimerie royale.

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

Feddersen, T., Gailmard, S., and Sandroni, A. (2009). Moral Bias in Large Elections: Theory and Experimental Evidence. *American Political Science Review*, 103(2):175–192.

Feddersen, T. and Pesendorfer, W. (1998). Convicting the Innocent: The Inferiority of Unanimous Jury Verdicts under Strategic Voting. *The American Political Science Review*, 92(1):23–35.

Fehrler, S. and Hughes, N. (2017). How Transparency Kills Information Aggregation: Theory and Experiment. *American Economic Journal: Microeconomics, Forthcoming*.

Fischbacher, U. (2007). z-tree: Zurich toolbox for ready-made economic experiments. *Experimental economics*, 10(2):171–178.

Fonseca, J. R. and Cardoso, M. G. (2007). Mixture-model cluster analysis using information theoretical criteria. *Intelligent Data Analysis*, 11(2):155–173.

Friedman, R. and Friedman, M. (1990). *Free to Choose*. A Personal Statement. Houghton Mifflin Harcourt.

Fudenberg, D., Rand, D. G., and Dreber, A. (2012). Slow to anger and fast to forgive: Cooperation in an uncertain world. *The American Economic Review*, 102(2):720–749.

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

Gerling, K., Grüner, H. P., Kiel, A., and Schulte, E. (2005). Information acquisition and decision making in committees: A survey. *European Journal of Political Economy*, 21(3):563–597.

Goeree, J. K. and Holt, C. A. (2004). A model of noisy introspection. *Games and Economic Behavior*, 46(2):365–382.

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

Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with orsee. *Journal of the Economic Science Association*, 1(1):114–125.

Guarnaschelli, S., McKelvey, R. D., and Palfrey, T. R. (2000). An experimental study of jury decision rules. *American Political Science Review*, 94(02):407–423.

Hall, M. (1992). Electoral Politics and Strategic Voting in State Supreme Courts. *The Journal of Politics*, 54(2):427–446.

Holmstrom, B. (1999). Managerial Incentive Problems: A Dynamic Perspective. *The Review of Economic Studies*, 66(1):169–182.

Huck, S. and Konrad, K. (2005). Moral Cost, Commitment, and Committee Size. *Journal of Institutional and Theoretical Economics*, 161:575–588.

Isaac, M. and Walker, J. (1988). Group Size Effects in Public Goods Provision: The Voluntary Contributions Mechanism. *The Quarterly Journal of Economics*, 103(1):179.

Ledyard, J. (1995). Public Goods: A Survey of Experimental Research. In Kagel, J. and Roth, A., editors, *The Handbook of Experimental Economics*. Princeton University Press, Princeton, NJ.

Levy, G. (2007). Decision Making in Committees: Transparency, Reputation, and Voting Rules. *The American Economic Review*, 97(1):150–168.

Li, H. (2001). A Theory of Conservatism. *Journal of Political Economy*, 109(3):617–636.

Li, H. and Suen, W. (2009). Viewpoint: Decision-making in committees. *Canadian Journal of Economics/Revue canadienne d'économique*, 42(2):359–392.

MacLachlan, G. and Peel, D. (2000). *Finite mixture models*. Wiley series in probability and statistics. Wiley.

Midjord, R., Rodríguez Barraquer, T., and Valasek, J. (2017). Voting in large committees with disesteem payoffs: A 'state of the art' model. *Games and Economic Behavior*, 104:430–443.

Morgan, J. and Várdy, F. (2012). Mixed Motives and the Optimal Size of Voting Bodies. *Journal of Political Economy*, 120(5):986–1026.

Palfrey, T. (2016). Experiments in Political Economy. In Roth, A. and Kagel, J., editors, *The Handbook of Experimental Economics*, pages 347–434. Princeton University Press.

Persico, N. (2004). Committee Design with Endogenous Information. *The Review of Economic Studies*, 71(1):165–191.

Powell, M. J. (2006). The NEWUOA software for unconstrained optimization without derivatives. In *Large-scale nonlinear optimization*, pages 255–297. Springer.

Rios, L. M. and Sahinidis, N. V. (2013). Derivative-free optimization: a review of algorithms and comparison of software implementations. *Journal of Global Optimization*, 56(3):1247–1293.

Stahl, D. O. and Wilson, P. W. (1994). Experimental evidence on players' models of other players. *Journal of economic behavior & organization*, 25(3):309–327.

Stahl, D. O. and Wilson, P. W. (1995). On players' models of other players: Theory and experimental evidence. *Games and Economic Behavior*, 10(1):218–254.

Tyran, J.-R. (2004). Voting when money and morals conflict: an experimental test of expressive voting. *Journal of Public Economics*, 88(7):1645–1664.

Visser, B. and Swank, O. H. (2007). On Committees of Experts. *The Quarterly Journal of Economics*, 122(1):337–372.

Weizsäcker, G. (2003). Ignoring the rationality of others: evidence from experimental normal-form games. *Games and Economic Behavior*, 44(1):145–171.

A Proofs for Section 3

Proof of Lemma 1: Given that $\tau_j(,,,) = 1$ for $j \neq i$, the committee outcome is independent of v_i, which implies that $E(p_i|v_i = R, M^{\#}, m_i, s_i) = E(p_i|v_i = B, M^{\#}, m_i, s_i) + K$. Therefore, $\tau_i(,,,) = 1$ is a best response. ■

Proof of Lemma 2: First, note that $piv(x)$ is equal to zero for $x = 0, 1$, and is maximized at $x = 0.5$. Moreover, $piv(x)$ is strictly increasing over the domain $[0, 0.5)$ and strictly decreasing over $(0.5, 1]$. Therefore, if $M^{\#} = S^{\#} < S^{piv}$, then the following expression holds for all $x \in [0, 1]$:

$$piv(x)[\Pr(\omega = B|S^{\#}) - \Pr(\omega = R|S^{\#})] \leq piv(0.5)[\Pr(\omega = B|S^{\#}) - \Pr(\omega = R|S^{\#})] < K.$$

This implies that $E(p_i|v_i = R, M^{\#}, m_i, s_i) > E(p_i|v_i = B, M^{\#}, m_i, s_i)$ for any $\tau(M^{\#})$ and that there is no equilibrium in which $\tau(,,,) < 1$. Second, if $M^{\#} > S^{piv}$, then there exists a unique $\tau^*(M^{\#}) \in [0, 0.5]$ such that the equilibrium condition, $piv(\tau^*(M^{\#}))[\Pr(\omega = B|S^{\#}) - \Pr(\omega = R|S^{\#})] = K$, holds. By the symmetry of $piv(x)$ about 0.5, $piv(1 - \tau^*(M^{\#}))[\Pr(\omega = B|S^{\#}) - \Pr(\omega = R|S^{\#})] = K$, and $1 - \tau^*(M^{\#})$ is the unique equilibrium in $[0.5, 1]$. ■

Proof of Lemma 3: For $M^{\#} < \bar{S}$, $o_i = B$ is a strictly dominated strategy since $\Pr(\omega = B|S^{\#}) - \Pr(\omega = R|S^{\#}) < K$. For $M^{\#} > \bar{S}$, however, $o_i = B$ is a best response to any $\tau < 1$ since $\Pr(\omega = B|S^{\#}) - \Pr(\omega = R|S^{\#}) > K$. ■

Proof of Lemma 4: Given $\tau(M^{\#}) = 1$, the committee outcome is independent of m_i, which implies that any messaging strategy is an equilibrium. ■

Proof of Proposition 1: We prove the result by contradiction. Assume an equilibrium exists with truthful messaging in the deliberation stage, and that all agents play strategy $\tau^*(M^{\#})$ in the voting stage. Note that the formulation of the proposition implies that given K, $\tau^*(M^{\#}) > 0$ for $M^{\#} = N$ (that is, $S^{piv} < N$, otherwise the only equilibrium is the babbling equilibrium).

The proof stems from the following two observations: (1) since $K > K'$ implies that $S^{piv} > (N+1)/2$, when $M^{\#} = (N+1)/2$, the unique equilibrium in the voting subgame is for all agents to vote for R; (2) as shown by equation 1, $\tau^*(M^{\#})$ is increasing in $M^{\#}$.

Now, consider the expected payoff of an agent, i', who has a signal of R, but who deviates to $m_{i'} = B$. Also, assume that i' plays strategy $\tau^*(M - 1)$ – that is, conditional on $S^{\#}$, i' continues to play the same strategy as under truthful communication. This implies that i''s expected expressive payoff is unchanged conditional on $S^{\#}$, and i''s expected payoffs change only due to the change in the probability that $X = B$ given $S^{\#}$. First, note that by (1), $\tau^*(M^{\#}) = 0$ for $M^{\#} \leq (N+1)/2$, $\Pr(X = B|S^{\#} < N/2, m_{i'} = B) = 0$. Second, by (2), $\Pr(X = B|S^{\#} > N/2, m_{i'} = B) \geq \Pr(X = B|S^{\#} > N/2, m_{i'} = R)$, and $\Pr(X = B|S^{\#} > N/2, m_{i'} = B) > \Pr(X = B|S^{\#} > N/2, m_{i'} = R)$ for $S^{\#} > S^{piv} - 1$.

Therefore, since i''s expected utility is strictly higher given an increase in $\Pr(X = B|S^{\#} > N/2)$, setting $m_{i'} = B$ is a best response. (Note, however, that the strategy $(\sigma(R) = \sigma(B) = 0, \tau^*(M^{\#} - 1))$ is not a best reply – given that other agents play the mixed strategy $\tau^*(S^{\#} - 1)$, i' has a best reply of voting R for all $M^{\#}$ $(\sigma(R) = \sigma(B) = 0, \tau(M^{\#}) = 1)$). ■

Proof of Proposition 2: Lemma 3 shows that given truthful messaging, it is an equilibrium for agents to unanimously submit an opinion for B in the deliberation stage if and only if selecting and voting for option B is optimal given the expressive cost. Next, we show that truthful messaging is an equilibrium given this voting strategy.

Consider the expected payoff of an agent, i', who has a signal of R, but who deviates to $m_{i'} = B$. Given $M^{\#} < \bar{S}$, all agents $i \neq i'$ will vote $v_i = R$ and i' has a best reply to also vote for R, which implies that the outcome is unchanged relative to truthful reporting. Given $M^{\#} \geq \bar{S}$, all agents $i \neq i'$ will vote $v_i = B$, which implies that i''s opinion is pivotal. This implies that for $M^{\#} > \bar{S}$, i' has a best reply of voting for B. Lastly, for $M^{\#} = \bar{S}$, i' has a best reply to vote for R, since $S^{\#} = M^{\#} - 1 < \bar{S}$.

Taken together, this shows that the committee outcome, X, and i''s behavior in the deliberation stage, conditional on $S^{\#}$, is unchanged by i''s deviation in the messaging stage. Therefore, i''s expected payoffs are not changed by the deviation to $m_{i'} = B$. Similarly, i' with $s_{i'} = B$ is made strictly worse off by deviating from truthful messaging, since the committee outcome will select R when $S^{\#} = \lceil \bar{S} \rceil$. This shows that truthful messaging and optimal information aggregation in the deliberation stage is an equilibrium under Consensus for all parameter values. ■

Sincere Voting Before comparing information aggregation under Majority and Consensus, we first consider whether there are equilibria under Majority that optimally aggregates information without truthful messaging. Note that under Consensus, truthful messaging is a necessary condition for optimal information aggregation since, to achieve efficiency, each agent needs to condition their vote on the aggregate signal profile. Under Majority, however, optimal information aggregation can be achieve without communication as long as each agent votes their signal ($v_i = s_i$). Therefore, we consider conditions under which $v_i = s_i$ is an equilibrium voting strategy.

First, we define K'' as follows:

$$K'' = \Pr(S^{\#} = 2 | s_i = B)[\Pr(\omega = B | S^{\#} = 2) - \Pr(\omega = R | S^{\#} = 2)].$$

The following lemma characterizes when perfectly informative voting is an equilibrium under Majority, given uninformative messaging.

Lemma 5 (Informative Voting: Majority)**.** *Babbling at the message stage, e.g.* $\sigma(s_i) = 1$ *for all i, and informative voting,* $v_i = s_i$*, at the voting stage is an equilibrium strategy if and only if* $K \leq K''$.

Proof: Assume i receives a blue signal, $s_i = B$. Since messaging is uninformative and since, given a strategy of $v_i = s_i$, i's vote is pivotal only when $S = 2$, i's expected relative payoff of voting B is equal to:

$$\Pr(S^{\#} = 2 | s_i = B)[\Pr(\omega = B | S^{\#} = 2) - \Pr(\omega = R | S^{\#} = 2)] - K,$$

which is positive iff $K < K''$. ■

Similar to Proposition 1, Lemma 5 shows that perfectly informative voting is only an equilibrium when K is small. Also, note that there are no equilibria with partially informative messaging and perfectly informative voting for $K > K''$: if messaging is partially informative, then there exists some $M^{\#}$ such that $\Pr(S^{\#} = 2|s_i = B, M) \leq \Pr(S^{\#} = 2|s_i = B)$, in which case it is not an equilibrium for i to vote B for all $M^{\#}$.

Proof of Theorem 1: First, note that Consensus limits the outcome set to points such that $v_i = X$. Therefore, given utility functions of the form $u : \{X, v_i, \omega\} \to \mathbb{R}$, agents have homogenous payoffs at all terminal nodes—this implies that there exists an equilibrium with truthful communication and coordination on the optimal outcome.

For clarity, however, we prove the result by contradiction. Assume there exists a deviation from the symmetric strategy $m_i = s_i$, and $o_i = R$ if $\mathbf{m} \in \mathbf{S}^R$ and $o_i = B$ if $\mathbf{m} \in \mathbf{S}^B$ that results in a strictly positive increase in i's expected payoffs.

Given that Consensus limits the outcome set to points such that $v_i = X$, any deviation that changes i's expected payoffs requires that the expected committee outcome changes as a function of the underlying vector of signals for at least one profile of signals, $\mathbf{s}' \in \mathbf{S}$. This implies that any profitable deviation results in either (1) $X = B$ with positive probability for some $\mathbf{s} \in \mathbf{S}^R$, (2) $X = R$ with positive probability for some $\mathbf{s} \in \mathbf{S}^B$, or both. However, by definition, either (1) or (2) imply a strict decease in i's expected payoffs, which is a contradiction. ■

Proof of Corollary 2: The result follows as a corollary to Lemma 2. Given truthful messaging, all agents are at the same information set at the voting stage. However, it is not an equilibrium for all agents to set $\tau(X', \mathbf{m} = \mathbf{s}') = 1$, since one agent deviating to $\tau(X', \mathbf{m} = \mathbf{s}') = 0$ does not change the committee outcome and $E(u(X', v_i \neq X', \omega)|\mathbf{s}') > E(u(X', v_i = X', \omega)|\mathbf{s}')$. ■

B Additional Material and Robustness Analyses

Additional graphs and tables Figure 3 provides a composite screen-shot that displays all queries and all pieces of information that were available to subjects at some point during the experiment. Table 6 describes the relative frequency and number of votes for B across all information sets in all four treatments.

Learning In this subsection we show that the qualitative results presented in above are robust to excluding the first 25 rounds of the experiment (learning). Figures 4 and 5 replicate the Figures refered to when discussing Questions 1 and 2 for the restricted data set comprising only the second halves of all sessions. The patterns are virtually indistinguishable. Table 7 replicates the probit regression refered to in the discussion of Question 3, including a dummy "Late" distinguishing whether subjects are in the first or second halves of their sessions. With only two exceptions, "Late" again is insignificant. The remaining statistical results relied upon in the discussion of Questions 3 and 4 in the main text distinguish first and second halves of sessions explicitly, thus establishing robustness to learning explicitly.

Multiple voting rounds under Consensus As we discuss in the main text, there is an asymmetry in our operationalization of Majority and Consensus. In particular, to replicate a consensus decision rule, we allowed for up to three rounds of voting in the Consensus treatment. While we do not have data on subjects' decisions if the status quo had been implemented following a non-unanimous first round of voting, we can consider this counterfactual by looking at the first-round votes. Figure 6 shows the committee outcome that would occur given the first-round votes and a counterfactual first-round Consensus rule (left graph), and a counterfactual first-round Majority rule (right graph).

Focusing on the case of $S = 3$, we see that information aggregation remains higher under Consensus under both counterfactual assumptions. Moreover, previous research has shown that under one-sided unanimity, subjects are less likely to exercise their veto and vote for the status quo (again, see Guarnaschelli et al., 2000 and Goeree and Yariv, 2011). If one-sided unanimity were to influence subject's behavior is a similar fashion in our experiment, then this would imply that the counterfactual first-round Consensus rule results are a lower bound for information aggregation (for the case of $S = 3$). Together, these results suggest that the finding of higher information aggregation under Consensus is not due to the multiple rounds of voting.

Figure 7 also includes the results for both counterfactual treatments for the second half of the experiment. Interestingly, the difference between the two counterfactual treatments remains high, suggesting that subject do not converge to a first-round consensus.

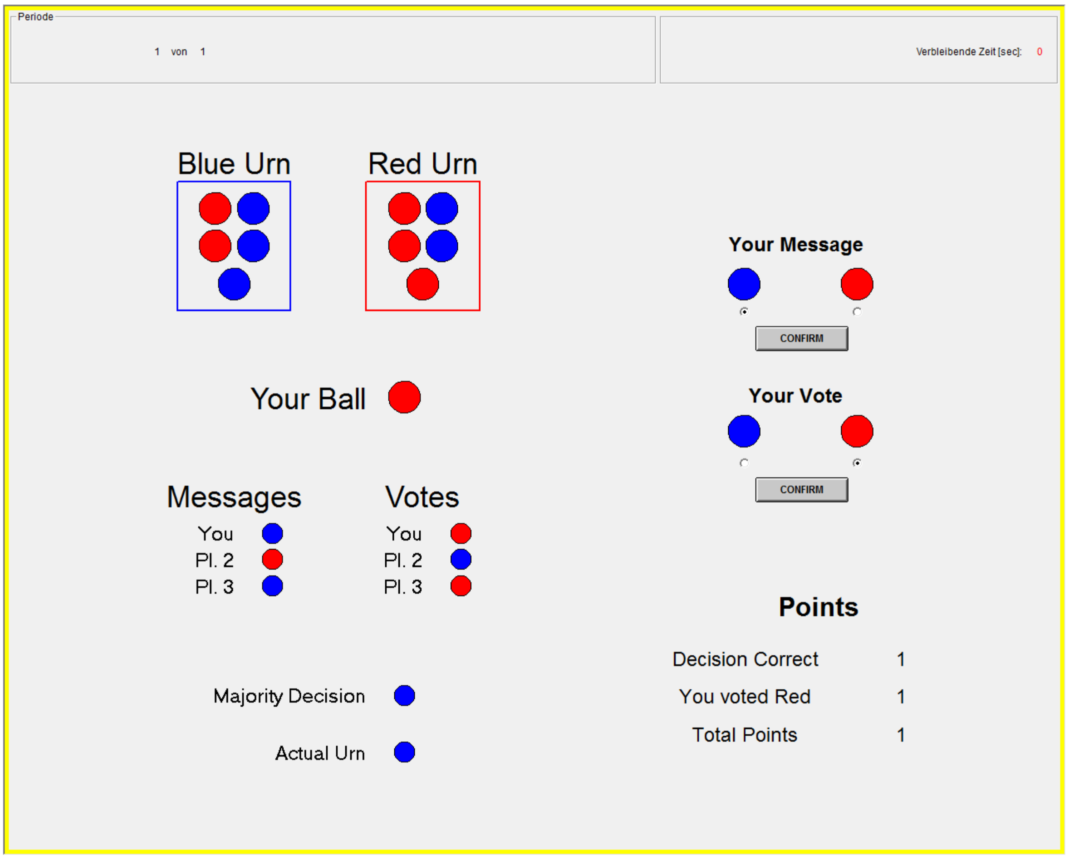
Finite mixture model: robustness checks Table 8 shows that eliminating components (strategy classes) from the analysis leads to worse values of the information criterion ICL-BIC, suggesting that no component should be eliminated.

Table 6: Proportion and number of votes for B

Treatment		High		Low	
M	s_i/m_i	Majority	Unanimity	Majority	Unanimity
0	red/red	0.02 (198)	0.00 (399)	0.02 (304)	0.00 (404)
0	blue/red	0.08 (24)	0.00 (63)	0.23 (53)	0.00 (43)
1	red/red	0.05 (464)	0.01 (487)	0.07 (523)	0.02 (511)
1	red/blue	0.09 (46)	0.00 (17)	0.14 (36)	0.12 (17)
1	blue/blue	0.08 (218)	0.02 (258)	0.08 (276)	0.01 (276)
1	blue/red	0.22 (64)	0.06 (63)	0.18 (101)	0.03 (75)
2	red/red	0.25 (228)	0.35 (240)	0.36 (237)	0.62 (192)
2	red/blue	0.06 (110)	0.19 (42)	0.32 (57)	0.47 (19)
2	blue/blue	0.40 (456)	0.41 (516)	0.56 (507)	0.63 (435)
2	blue/red	0.31 (55)	0.62 (39)	0.49 (45)	0.66 (35)
3	red/blue	0.16 (85)	0.54 (13)	0.45 (20)	0.83 (12)
3	blue/blue	0.69 (302)	0.93 (263)	0.69 (241)	0.96 (231)

Proportion of votes for B as a function of the aggregate message profile (M) and the individual signal/message (number of observations are reported in parentheses). We use first-round votes for the unanimity treatments.

Figure 3: Composite screenshot (translated; the original version is provided in the supplementary material)



Note: This screenshot simultaneously displays all queries and all pieces of information that were available at some point during the experiment. All items are in the positions they had been displayed, and they were displayed in the following order.

1. Show urns and drawn ball (displayed for the entire game)
   Shows the two jars ("Blue Urn" and "Red Urn") and the ball drawn ("Your ball"). These items remain on the screen for the entire game.

2. After five seconds, query for message (no time limit)
   Now the box "Your Message" appears with the two balls underneath to choose from. Subjects submit the message by clicking "OK", there is no time limit. Once the message is submitted, the box disappears.

3. When all messages are submitted, they are displayed (displayed for the remainder of the game)
   Now the box "Messages" on the left appears, displaying the messages of all three subjects. These items remain on the screen for the rest of the game.

4. After five seconds, query for vote (no time limit)
   Now the box "Your Vote" appears with the two options to choose from. Subjects submit their vote by clicking "OK", there is no time limit. Once the vote is submitted, the box disappears.

5. When all votes are submitted, they are displayed (displayed for the remainder of the game)
   Now the box "Votes" on the left appears, displaying the votes of all three subjects. These items remain on the screen for the rest of the game (in Majority or in Consensus if decision unanimous or the third vote was taken) or disappear (in Consensus otherwise, where the voting stage is restarted).

6. After five seconds, the decision taken by the committee ("Majority Decision"), the urn originally chosen by Nature ("Actual Urn") and the payoff information is displayed ("Points"). The majority decision and numbers displayed here are entirely artifical. The information remains on the screen for 10 seconds, after which a new game starts.

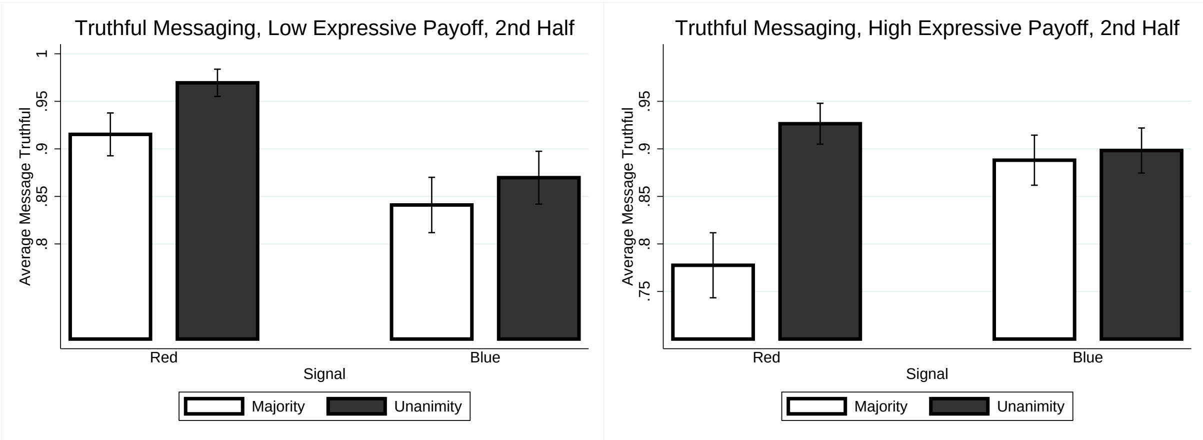


Figure 4: Truthful reporting for rounds 26 – 50.

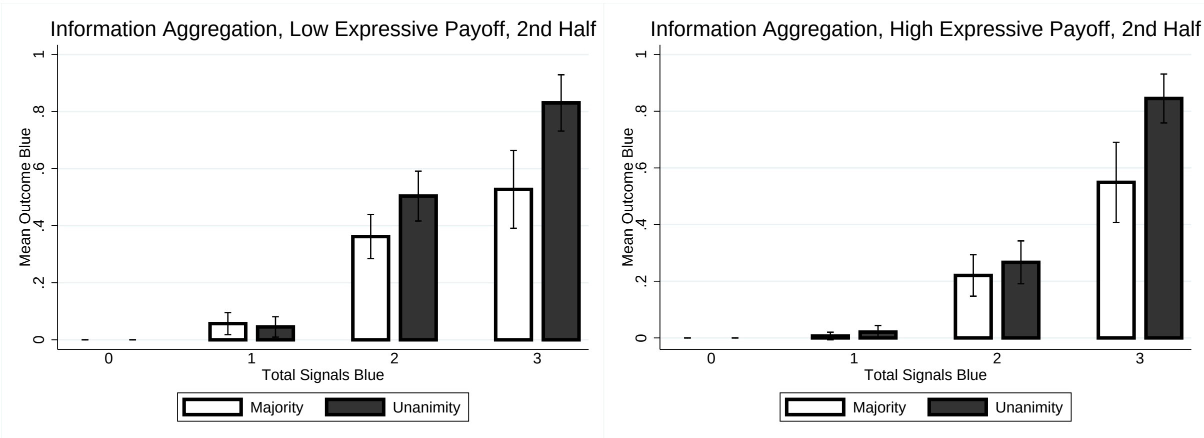


Figure 5: Information aggregation for rounds 26 – 50.

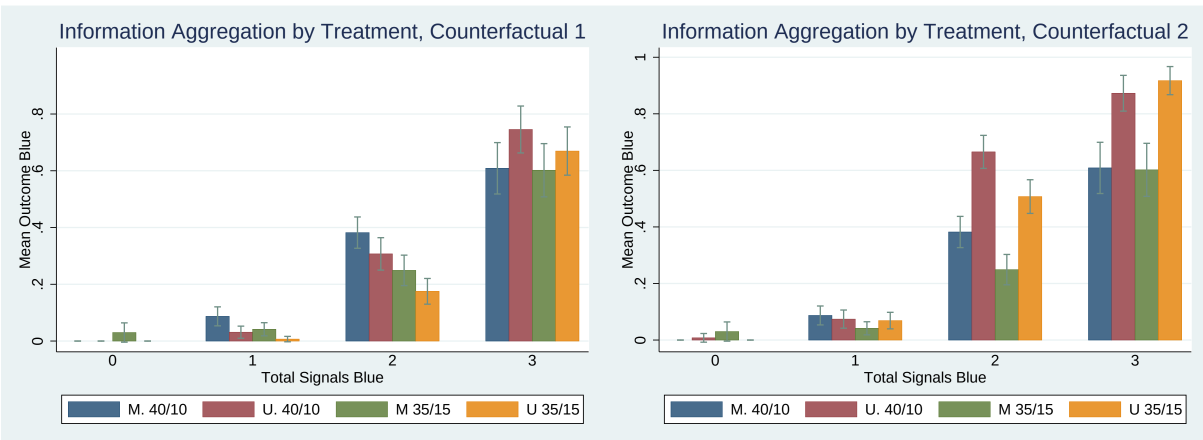


Figure 6: Average levels of outcome *Blue*, assuming a counterfactual first-round Consensus rule (left graph) and a counterfactual first-round Majority rule (right graph), as a function of the aggregate profile of signals.

Table 7: Probit estimations to explore subject learning

	(1)	(2)	(3)
Vote *Blue*	Low	High	Joint
Own signal *B*	1.477***	1.627***	1.544***
	(0.168)	(0.158)	(0.111)
Number of other's messages *B*	1.692***	1.622***	1.656***
	(0.136)	(0.129)	(0.0928)
Majority	1.372***	1.357***	1.329***
	(0.284)	(0.330)	(0.209)
Majority*own signal	-0.510**	-0.513*	-0.493***
	(0.198)	(0.216)	(0.140)
Majority*other's messages	-0.902***	-0.905***	-0.890***
	(0.162)	(0.161)	(0.112)
Late	-0.250	-0.611**	-0.341*
	(0.252)	(0.228)	(0.166)
Late*own signal	0.180	0.0851	0.105
	(0.132)	(0.141)	(0.0939)
Late*other's messages	0.215	0.168	0.128
	(0.173)	(0.144)	(0.110)
Late*majority	-0.110	0.0482	-0.0738
	(0.249)	(0.266)	(0.175)
Late*majority*other's messages	-0.147	-0.00879	-0.0322
	(0.191)	(0.179)	(0.127)
High			-0.322***
			(0.0833)
Constant	-3.264***	-3.447***	-3.186***
	(0.240)	(0.258)	(0.168)
N	4650	4650	9300

Subject-level clustered standard errors in parentheses, "late" indicates round > 25.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

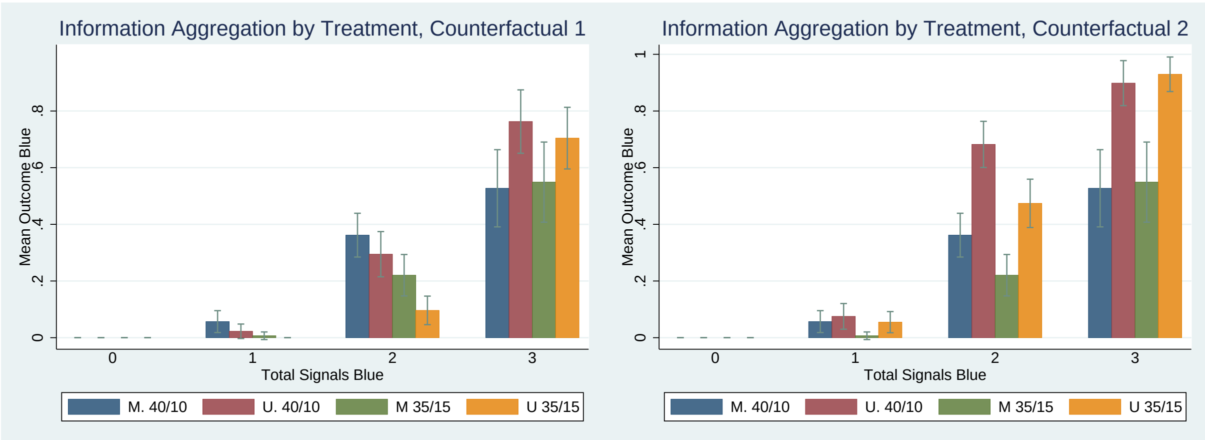


Figure 7: Average levels of outcome *Blue*, assuming a counterfactual first-round Consensus rule (left graph) and a counterfactual first-round Majority rule (right graph), for rounds $26-50$.

Table 8: Robustness check on estimated strategy weights, testing whether all strategy classes have significant weight. The test based on ICL-BIC (less is better), and we find that no strategy class may be eliminated without increasing ICL-BIC. Format is equal to Table 5b

	Strategy weights in population					Strategy parameters					
	Noise	Honest	StratRed	StratBlue	FreeRide	ε	π_{Lie}	π_{High}	π_{Med}	π_{Low}	ICL-BIC
All games per session											
Majority 35 15	0.12	0.44	0.19	0	0.26	0.04	0.42	0.73	0.35	0.08	6368.79
Majority 40 10	0.07	0.45	0.25	0.03	0.2						
Unanimity 35 15	0.07	0.78	0.11	0.04	0						
Unanimity 40 10	0.04	0.75	0.18	0.02	0						
Majority 35 15	0.12		0.2	0.18	0.51	0.05	0.14	0.62	0.23	0.04	6746.9
Majority 40 10	0.11		0.25	0.22	0.42						
Unanimity 35 15	0.1		0.13	0.7	0.06						
Unanimity 40 10	0.06		0.21	0.53	0.19						
Majority 35 15	0.16	0.21		0.05	0.58	0.05	0.13	0.75	0.25	0.05	6770.39
Majority 40 10	0.18	0.25		0.05	0.53						
Unanimity 35 15	0.13	0.8		0	0.07						
Unanimity 40 10	0.09	0.52		0.11	0.27						
Majority 35 15	0.12	0.44	0.19		0.26	0.04	0.42	0.72	0.35	0.08	6368.55
Majority 40 10	0.07	0.48	0.25		0.21						
Unanimity 35 15	0.11	0.78	0.11		0						
Unanimity 40 10	0.04	0.78	0.18		0						
Majority 35 15	0.13	0.57	0.13	0.16		0.04	0.54	0.5	0.44	0.17	6468.46
Majority 40 10	0.07	0.62	0.11	0.2							
Unanimity 35 15	0.07	0.81	0.08	0.04							
Unanimity 40 10	0.04	0.79	0.15	0.02							
Majority 35 15	0.13		0.22		0.64	0.05	0.13	0.35	0.19	0.25	6948.35
Majority 40 10	0.11		0.32		0.57						
Unanimity 35 15	0.1		0.13		0.76						
Unanimity 40 10	0.06		0.27		0.67						
Majority 35 15	0.16	0.25			0.59	0.05	0.13	0.74	0.25	0.05	6766.47
Majority 40 10	0.18	0.3			0.53						
Unanimity 35 15	0.13	0.8			0.07						
Unanimity 40 10	0.09	0.64			0.27						
Majority 35 15	0.23	0.66	0.11			0.05	0.67	0.68	0.43	0.15	6677.39
Majority 40 10	0.16	0.75	0.09								
Unanimity 35 15	0.11	0.82	0.07								
Unanimity 40 10	0.04	0.82	0.14								
Majority	0.09	0.44	0.22	0.01	0.23	0.04	0.41	0.72	0.35	0.08	6345.24
Unanimity	0.06	0.77	0.15	0.03	0						
Majority	0.11		0.23	0.19	0.47	0.05	0.14	0.62	0.22	0.03	6728.42
Unanimity	0.08		0.17	0.61	0.14						
Majority	0.17	0.23		0.05	0.55	0.05	0.13	0.75	0.25	0.05	6758.34
Unanimity	0.11	0.66		0.06	0.17						
Majority	0.09	0.46	0.22		0.23	0.04	0.42	0.72	0.35	0.08	6349.6
Unanimity	0.08	0.78	0.15		0						
Majority	0.1	0.6	0.12	0.18		0.04	0.54	0.5	0.44	0.17	6448.95
Unanimity	0.06	0.8	0.12	0.03							
Majority	0.12		0.27		0.61	0.05	0.13	0.35	0.19	0.25	6934.63
Unanimity	0.08		0.2		0.72						
Majority	0.17	0.27			0.56	0.05	0.13	0.75	0.25	0.05	6754.68
Unanimity	0.11	0.72			0.17						
Majority	0.19	0.71	0.1			0.05	0.67	0.67	0.43	0.15	6663.68
Unanimity	0.08	0.82	0.1								