Strategic Communication in Committees with Mixed Motives *

Yves Breitmoser[†] Justin Valasek[‡]

February 12, 2021

Abstract

This paper explores strategic communication in settings where committee members are held accountable, formally or informally, for their individual voting decisions. In a controlled laboratory experiment, we show that if decisions are made via majority, expressive payoffs introduce a free-rider problem that causes agents to communicate strategically, which prevents the committee from taking optimal decisions. In contrast, if decisions are made by unanimity, free-riding is mitigated since all agents are responsible for the committee's decision: under unanimity subjects are more truthful, respond more to others' messages, and are ultimately more likely to take the optimal decision.

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

JEL Classification Codes: D71, D72, C90.

*Thanks to Marina Agranov, Nageeb Ali, Inacio Bo, Agustin Casas, Chang-Koo Chi, Guillaume Fréchette, Dan Friedman, Steffen Huck, Navin Kartik, Aniol Llorente-Saguer, Andrea Mattozzi, Krisztina Molnar, Rebecca Morton, Salvatore Nunnari, Tom Palfrey, Karl Schlag, Andy Schotter, Sebastian Schweighofer-Kodritsch, Peter Norman-Sorensen, and Arturas Rozenas as well as seminar participants at the Helsinki School of Economics, University of Bristol, Norwegian School of Economics, University of Vienna, Southern Denmark University, University of Mannheim, NYU CESS Conference, CESifo PSE Conference and Rotterdam PE Workshop 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 "The Value of Consensus: Information Aggregation in Committees with expressive Payoffs." Financial support of the WZB Berlin and the DFG (project BR 4648/1 and CRC TRR 190) is greatly appreciated. Corresponding author: Justin Valasek.

†Bielefeld University. Contact e-mail: yves.breitmoser@uni-bielefeld.de

‡Norwegian School of Economics (NHH) and CESifo. Contact e-mail: justin.valasek@nhh.no

1 Introduction

Committees are a ubiquitous institution for making institutional 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, board members may benefit from supporting successful ventures, politicians may wish to signal their ideological position to the electorate, and FDA experts may seek to avoid blame for approving a drug that proves to have severe side-effects.[1]

In cases where committee members have mixed motives, 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 expressive biases (see for example Brennan and Buchanan, 1984, Callander, 2008 and Morgan and Várdy, 2012). However, an important factor to consider when assessing the impact of mixed motives on committee behavior is that voting in small and medium-sized committees is often preceded by open discussion. Deliberation has been shown to significantly improve a committee's ability to aggregate information (see Coughlan, 2000, Guarnaschelli et al., 2000), but the impact of mixed motives on the effectiveness of pre-vote deliberation is an understudied topic.[2]

[1]For example, David Willman of the LA Times later won a Pulitzer Prize in part for his article analyzing the FDA committee's decision to approve Posicor, was withdrawn after being linked with over 100 deaths, by a 5-3 vote (Willman, 2000). His article explicitly names two of the committee members who voted against the drug, and cites a third committee member who could not participate in the vote due to financial conflict as stating "You do wonder how the world would perceive it. I'm glad I didn't vote…"

[2]Pre-vote communication has been widely studied in the case where committee members have conflicting prefer-

In this paper, we explore the effect of expressive payoffs—we use the term expressive payoff to refer to payoffs that depend on the individual's vote—on communication and information aggregation in committees. Additionally, we explore how the effect of expressive payoffs on communication may be sensitive to the decision rule used by the committee. In particular, we highlight and test the novel prediction that when committees are subject to mixed motives and have access to communication, a majority rule results in strategic communication while a unanimity rule will lead to truthful communication and information aggregation.

We begin by illustrating that under majority rule, expressive payoffs give agents an incentive to convince other's to support a given option, while individually voting for the opposite option. To outline the intuition, take the case of the European Stability Mechanism (ESM), which is composed of the finance ministers of Eurozone countries and takes decisions regarding the management and disbursement of bail-out funds to individual Eurozone countries via a vote. While the finance ministers may want to approve bailouts that are necessary to secure financial stability, bailouts of other EU countries are not always popular with the politicians' domestic audiences and their vote may therefore impact their chances of reelection. Accordingly, suppose that the finance ministers face a reelection cost for voting to approve a bailout independent of whether a bailout is necessary to secure the financial stability of the Eurozone, and that the decision to approve a bailout is taken by majority rule.[3]

Additionally, assume each minister receives an independent signal that is informative of whether or not a bailout is necessary, and that the ministers privately deliberate prior to voting (EU finance ministers deliberate behind closed doors before publicly voting whether to approve a bailout). In this case, even assuming that all ministers truthfully reveal their private information during deliberation, it will not be part of an equilibrium strategy for all ministers to vote to approve the bailout when deliberation indicates (i.e. the shared signals indicate) that this is the optimal decision—if all other ministers members vote to approve, a single minister can vote to

ences over the *committee* decision (see for example Li, 2001, Austen-Smith and Feddersen, 2006 and Goeree and Yariv, 2011). In contrast, to the best of our knowledge, communication and mixed motives has not been addressed.

[3] Decisions in the ESM are currently taken using a unanimity rule, recent proposals to replace the ESM with a European Monetary Fund (EMF) have discussed using a majority rule (see European Commission, 2017; Sapir and Schoenmaker, 2017).

reject the bailout without changing the committee outcome, and hence will avoid the reelection cost. Therefore, given truthful communication it will be a joint best reply for the ministers to mix over voting to approve and reject, which introduces a probability that the ESM votes against a bailout even when the evidence indicates that a bailout is necessary to secure financial stability.

In turn, this collective-action problem at the voting stage affects the value of communication. That is, consider the position of a minister who has access to an insight that suggests that a bailout may be unnecessary. This minister knows that, relative to the appropriate threshold of doubt, the committee will be less likely to approve a bailout due to the bias introduced in the voting stage. Should the minister reveal their insight to the other ministers during deliberation? Anticipating the bias towards not approving the bailout, the minister actually has an incentive to withhold their insight, and thereby move the expected decision closer to optimal threshold for approval. Due to this incentive to strategically withhold or misrepresent information, truthful communication will fail to be part of an equilibrium strategy under majority rule.

We test this prediction experimentally and find that under majority subjects systematically misreport their messages in a pre-vote round of binary cheap talk in a direction consistent with intuition outlined above. Additionally, 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.

Next, we consider behavior under the same expressive payoffs when the committee uses a unanimity rule to reach a decision. Continuing with the example above, requiring unanimity implies that all ministers must vote approve for a bailout to be approved. This removes the collective action problem in the voting stage— all ministers must forgo the expressive payoffs for a bailout to be approved. Therefore it is an equilibrium strategy for all ministers to vote to approve the bailout when deliberation indicates that this is the optimal decision. In turn, since that the committee votes optimally given the information, ministers no longer have an incentive to communicate strategically under unanimity. That is, in contrast with majority, under a unanimity rule, an equilibrium exists with truthful communication and committee-optimal decisions.

Experimentally, we find that with expressive payoffs subjects are indeed more truthful under unanimity relative to majority, and as predicted, unanimity 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). Additionally, we find that no subjects are classified as pursuing a "free-rider" strategy when the committee takes decisions via unanimity, suggesting that unanimity avoids the collective action problem that occurs under majority rule when the committee is subject to mixed motives.

As mentioned above, our paper's main contribution is to the literature that has explored the impact of mixed motives on committee behavior (Callander, 2008 and Morgan and Várdy, 2012)) and the literature that has focused on the efficacy of communication when agents have conflicting preferences over the committee decision (see Li, 2001, and Austen-Smith and Feddersen, 2006, Goeree and Yariv, 2011). In this case, agents have an incentive to deviate from truthful communication in an attempt to bias the committee decision towards their preferred option—that is, opposite from our prediction, agents have an incentive to misreport for the same option that they individually vote for. Importantly, heterogeneous preferences over the committee outcome also lead to distorted committee decisions under both majority and unanimity (Goeree and Yariv, 2011).[4]

In contrast, the mechanism we highlight here does not rely on heterogeneous preferences over the committee decision—all agents receive the same instrumental benefits if the committee makes the correct decision. However, due to expressive payoffs, agents have strategic incentive

[4]Interestingly, Goeree and Yariv (2011) do not find that subjects communicate strategically in this setting, while we find strong evidence for strategic communication.

to misreport to persuade others to vote for the option that is collectively optimal, but individually costly. Moreover, this collective-action problem only occurs under majority, since majority admits the possibility that the committee selects the collectively optimal option even when a minority "free-rides" and votes for the individually optimal decision. In contrast, Unanimity rule holds all members equally responsible for the committee decision, which facilitates committee-optimal behavior.

Our paper is also related to the literature on reputation payoffs in committees, which considers the Holmstrom (1999) model of career concerns applied to a committee setting. Here, agents do not receive payoffs related to the committee decision—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, 2018, for a review).[5] Visser and Swank (2007), however, consider the different decision rules in a setting with communication and reputation payoffs (additionally, Levy, 2007, compares majority and unanimity 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 decision.

Our paper proceeds as follows: Section 2 introduces the theoretical model and presents our formal results. Section 3 describes our experiment testing the predictions of the model and Section 4 presents the analysis of the experimental results. Section 5 concludes. The formal proofs, the experimental instructions and a number of robustness checks are provided in the Appendix.

[5] Fehrler and Hughes (2018) 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 principal.

2 The Model

We begin by considering a Condorcet setting with pre-vote communication (cheap talk), binary signals and a simple expressive payoff for voting for one of the two options. This setup allows us to clearly illustrate the main properties of voting in the presence of expressive payoffs—in Section A in the Online Appendix we extend the model to show that these predictions are robust to a generalized signaling space and expressive payoffs that also depend on the state of the world and the committee decision. 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 committee decisions.

2.1 Basic Framework

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.

The underlying state of the world is denoted by $\omega \in \Omega = \{R,B\}$. Agents do not observe the state of the world, and all agents have a common prior over the state of the world, denoted $P_R = \Pr(\omega = R)$, that specifies that each state is equally likely ($P_R = 1/2$). Additionally, each committee member receives a private signal from a binary signal space S, $s_i \in \{R,B\}$, with $\Pr(s_i = x|\omega = x)$ equal to $\alpha \in (1/2, 1)$ for any $x \in \{R,B\}$. For the common-value component of payoffs, each agent receives a payoff of 1 if the committee chooses the option that matches the underlying state of the world, and a payoff of 0 otherwise; for the expressive component of payoffs, each agent has a simple expressive voting bias and receives a payoff of $K \in (0,1)$ conditional on voting for option R. That is:

$$u^1(X,\omega) + u^2(X,\omega,v_i) = I_{X=\omega} + K * I_{v_i=R},$$

where $I_{v_i=R}$ takes a value of 1 if $v_i = R$ and 0 otherwise. This simple voting bias is independent of the state of the world and the decision of the committee—however, we generalize our findings to more complex vote-contingent expressive payoffs in section A of the Appendix.

Given the symmetric structure of the model, a positive payoff of K for voting for R is equivalent to a negative payoff of $-K$ for voting for B. Therefore, the model captures both positive and negative expressive payoffs. Also, since the signal space is binary, the profile of signals and messages can be characterized by the number of signals and messages, respectively, of B. We denote the aggregate number of signals of B by $S^{\#}$ ($S^{\#} = \sum_i I_{s_i=B}$), and the aggregate number of messages of B by $M^{\#}$ ($M^{\#} = \sum_i I_{m_i=B}$).

In our baseline model, we consider a committee that uses a majority rule; we introduce unanimity rule in the following subsection. Take d^v to be the decision rule used by the committee. We use the standard definition of a majority decision rule:

Definition 1 (Majority)**.** *Each player submits a vote $v_i \in \{R, B\}$. The final decision X is the option that receives a larger number of votes.*

The timing of the game is as follows:

1. Nature draws state $\omega \in \{R, B\}$ and sends private signals (s_i).
2. Committee members observe s_i and simultaneously send messages $m_i \in S$.
3. Committee members observe $M^{\#}$ and simultaneously submit votes $\in \{R, B\}$.
4. The committee decision is taken, the state is revealed and payoffs accrue.

The equilibrium concept we consider is symmetric Perfect Bayesian Nash (a formal definition is provided in the Appendix). By symmetry, we require that agents with the same signal who also sent the same message vote R with the same probability. Our theoretical results establish (non-) existence and properties of equilibria sustaining truthful communication. That is, all messages are sent with positive probably and all information sets relevant for defining the (stationary) strategies of our agents are on the path of play. Beliefs are therefore uniquely specified by Bayes' Rule and need not be specified explicitly. Given the beliefs consistent with Bayes'

Rule, expected utilities are also well-defined, and equilibria are defined straightforwardly.

We denote the game by $\hat{\Gamma} = \langle P_R, N, u, d^v \rangle$. Agents' strategies are pairs (σ_i, τ_i), where:

- $\sigma_i(m_i|s_i)$ is the probability of sending message $m_i \in S$ after receiving signal s_i,
- $\tau_i(s_i, m_i, M^{\#})$ is the probability of vote R after signal s_i, own message m_i, and the aggregate message profile.[6]

Definition 2 (Optimal Information Aggregation)**.** *We define* (Committee-) Optimal Information Aggregation *as replicating the decision taken by a single decision-maker (DM) with preferences $u(X, \omega, v^{DM}) = u(X, \omega, v_i)$, who has access to the complete profile of signals, and whose vote determines the committee decision, $X = v^{DM}$.*

We refer to the decision that optimally aggregates information as the *committee-optimal decision*. We also distinguish between optimal information aggregation and *information aggregation*: information aggregation implies that the committee selects the option that is most likely to coincide with the state of the world given the set of private signals; (committee-) optimal information aggregation, however, takes into account the expressive payoffs, which implies that, say, 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, and the payoffs from matching the decision to the state of the world are symmetric, social welfare may be maximized by information aggregation since the committee members' expressive payoffs may be small relative to the social impact of choosing the more likely option. Therefore, while we focus on optimal information aggregation in the theoretical analysis, we will consider both benchmarks when interpreting our experimental results.

[6]Since we consider stationary strategies and focus on the existence of an equilibrium with truthful communication under Unanimity, we omit a t subscript on τ_i without loss of generality.

2.2 Analysis

We begin by characterizing a well-known set of equilibria, existent in most games of information aggregation through voting, that do feature truthful communication in the communication stage: regardless of messages, all agents voting R is a mutual best response under Majority. Clearly, any message strategy can be sustained in these "non-responsive equilibria," including truthful communication.

Proposition 1 (Non-Responsive Equilibrium). *In any game $\hat{\Gamma}$ with d^v = Majority, there is a symmetric equilibrium where all agents truthfully communicate ($\sigma(s_i|s_i) = 1$) and vote for option R for any signal, message and profile of messages ($\tau(s_i, m_i, M^{\#}) = 1$).*

The rationale for Proposition 1 is that when $\tau(s_i, m_i, M^{\#}) = 1$ for all $i \neq j$, then j's vote cannot be pivotal, which implies that $v_j = R$ is a best response. Given $K > 0$ there is no non-responsive equilibrium with $\tau(s_i, m_i, M^{\#}) = 0$ under Majority, since if i's vote is not pivotal, then i has a best response of voting for R and receiving the expressive payoff. In the remainder of this section, we analyze equilibria with *Responsive Voting*, which we define as equilibria where $\tau(s_i, m_i, M^{\#})$ is not constant.

Responsive voting under Majority Next, we consider the existence of equilibria with truthful communication and responsive voting under Majority. In the analysis, three qualitatively similar but technically distinct cases need to be considered. We consider all cases in detail in the appendix, but for ease of exposition here, we focus on voting behavior that constitutes a mutual best response given truthful communication under the restrictions that (i) voting behavior conditions only on $M^{\#}$ and (ii) among the set of mutual best responses it maximizes the probability that the committee chooses the option with the majority of private signals. We denote such voting strategies as τ^M. In the other cases, agents may condition their voting behavior on their individual message, they might simply vote all R for some $M^{\#}$, or they might vote B with low probability, which yields results that are qualitatively similar for all classes of equilibria. Given our focus here, we simplify the notation of $\tau(s_i, m_i, M^{\#})$ to $\tau^M(M^{\#})$ for the remainder of this subsection.

The following lemma characterizes voting strategies in potential equilibria sustaining truthful communication and responsive voting satisfying this restriction.

Lemma 1 (Voting Stage: Majority)**.** *In any game $\hat{\Gamma}$ with d^v = Majority, for any symmetric equilibrium that exhibits truthful communication, $\sigma(s_i|s_i) = 1$, and responsive voting strategies $\tau^M = \tau^*(M^\#)$, there exists $S' \geq (N+1)/2$ such that*

- $\tau^*(M^\#) = 1$ *if* $M^\# < S'$*,*
- $\tau^*(M^\#) \in (0,1)$ *if* $M^\# \geq S'$*.*

We characterize S' and $\tau^*(M^\#)$ in the proof of Lemma 1 in the Appendix. Intuitively, a necessary condition for agents to vote for B with positive probability is that $M^\#$ is high enough such that the expected common-value payoff of selecting B is larger than the expressive payoff, K. However, this is not a sufficient condition; $\tau^*(M^\#) < 1$ can only be a symmetric best response if the expected common-value payoff of selecting B *times the probability of being pivotal*, given that other agents play $\tau^*(M^\#)$, is equal to the expressive payoff. Thus, S' is the minimum value of $M^\#$ such that this condition is satisfied for some $\tau^M(M^\#) \in (0,1)$.

More generally, Lemma 1 states that when R is committee optimal, all committee members vote for R with probability one in any truthful, responsive equilibrium. However, there is no truthful, responsive equilibrium in which all agents vote for B when B is committee optimal. Due to the free-rider problem highlighted in Theorem 1—each committee member would prefer that a majority vote for B, but to individually belong to the minority of agents that vote for R—all potential responsive symmetric equilibria involve mixing in the voting stage. This mixing over v_i introduces an aggregate bias toward option R, since there is a positive probability that the committee will select R even when B is committee optimal.

Lemma 1 characterizes the potential form of responsive equilibria with truthful communication under Majority. However, the result does not imply that such an equilibrium always exists. Based on Lemma 1, we next show that the existence of an equilibrium with truthful communication and responsive voting can be ruled out if the expressive payoff K is above a certain threshold.

Proposition 2 (Truthful Messaging under Majority)**.** *In any game* $\hat{\Gamma}$ *with* $d^v = Majority$*, there exists* $K' > 0$ *such that equilibria exhibiting truthful communication and responsive voting do not exist if* $K > K'$.

We explicitly characterize K' in the proof of Proposition 2 in the Appendix, but essentially, at K' it is no longer optimal to vote B if there are just $(N+1)/2$ signals for B. As a result, K' is fairly low even in small committees, as will become clear with our experimental design.

Proposition 2 shows that the free-rider problem in the voting stage has a knock-on effect on the messaging stage, eliminating equilibria with truthful communication and responsive voting for sufficiently high values of K. The intuition for this result is as follows: Independent of their own vote, each agent would prefer that the committee select option B when there are more signals for B (information aggregation). However, as detailed in Lemma 1, given truthful communication the expressive payoff biases the committee's decision toward R, since agents play mixed strategies at the voting game for $M^\# \geq S'$. Deviating from truthful communication and messaging B following a signal for R allows agents to reduce this bias since the probability that each agent votes for B is increasing in $M^\#$.

Because of this, agents face a tradeoff when considering a deviation to messaging B following a signal for R: on one hand, this decreases the committee's voting bias towards R; on the other hand, it may imply that the committee selects B when R has a higher number of signals. This tradeoff implies that truthful messaging and responsive voting can be an equilibrium for low levels of K, despite the committee's voting bias towards R. For large enough K, however, deviating to messaging B (and voting R) following a signal for R is individually optimal, thereby eliminating equilibria with truthful communication and responsive voting. We shall refer to this strategy, falsely reporting B while personally voting R, as the "free-riding" strategy.

2.2.1 Information aggregation under Unanimity

In this subsection, we consider committee-decisions under a unanimity rule. That is, we consider a decision rule that requires unanimous agreement: the committee reaches a decision only when all members of the committee vote for the same option. This form of unanimity rule is consistent with the legal definition of unanimity where, e.g., for a decision to be reached either all jury members must vote "guilty" or all members must vote "not guilty."[7] Accordingly, we model unanimity as a repeated series of informal straw polls, where a final decision is reached only when the straw poll is unanimous. This contrasts with a status-quo unanimity rule, common in international organizations, where one option is designated as a status-quo and unanimous agreement is required for any change from the status quo.[8] However, in the Appendix we show that a status-quo unanimity rule is partially robust to expressive payoffs, while a unanimity rule that requires unanimous agreement for any decision eliminates the free-rider problem for any expressive payoff.

We model the process by which the committee coordinates on a unanimous decision as a sequence of (internal) straw polls. Each straw poll results in a committee decision, $X \in \{R,B,D\}$, where decision "D" denotes "disagreement" in case unanimity is not achieved.:

Definition 3 (Unanimity)**.** *The straw poll proceeds in rounds, and each round each player submits a vote $v_i^t \in \{R,B\}$, $t \in \{0,1,\dots\}$. If the straw poll is unanimous then voting stops, each agent's final vote v_i is equal to their vote in the straw poll, and the committee decision is equal to the unanimous decision. Otherwise, a new round of the straw poll begins. In case of perpetual disagreement, $X = D$.*

[7] See American Bar Association (2013) for detailed description of jury deliberations. In federal jury trials in the US, if a unanimous decision cannot be reached after extensive jury deliberation, a "hung jury" results and a mistrial is declared. A mistrial, however, is not equivalent to an acquittal. As described by the American Bar Association (2013), a hung jury implies that: "The case is not decided, and it may be tried again at a later date before a new jury."

[8] In the formal literature, unanimity has been most commonly modeled as a status-quo unanimity rule (see Feddersen and Pesendorfer, 1998 and Maggi and Morelli, 2006). For a notable exception, see Coughlan (2000).

For simplicity, we focus on stationary strategies in the voting stage. While we do not include a discount term, including discounting between intermediate rounds of voting does not qualitatively impact our results. As we show later, an equilibrium of the game exists in which agents play stationary strategies and reach a decision in the first round of voting, so theoretically the straw poll is not necessary for agents to exchange information. Also, in our baseline model expressive payoffs are contingent on the final vote (v_i) only; we consider the case of expressive payoffs in the communication and coordination (straw-poll voting) stages in the Appendix. Lastly, in case of perpetual disagreement, $X = D$, all payoffs are zero ($u(D, \omega, v_i) = 0$).

The procedure for the unanimity rule specified above is just one of many ways to model the process by which the committee coordinate on a unanimous decision. We illustrate our results using this particular procedure since it follows the general process often used in committees: agents first publicly share their private signals, then "deliberate" using a sequence of straw polls, and reach a final decision when the outcome of a straw poll is unanimous (see Guarnaschelli et al., 2000; Goeree and Yariv, 2011). Note that the feature of unanimity rule that is most important for our analysis is that for a given decision to be chosen, all agents must unanimously vote for that decision.

Next, we clarify that the existence of an equilibrium with optimal information aggregation under Unanimity obtains in any game $\hat{\Gamma}$. We establish this result using the same line of reasoning that we use to derive the non-existence result under Majority. That is, first consider the case in which the committee truthfully communicates; the following lemma characterizes voting behavior in equilibria the feature truthful communication.

Lemma 2 (Voting Stage: Unanimity)**.** *In any game $\hat{\Gamma}$ with d^v = Unanimity, for any symmetric equilibrium that exhibits truthful communication, $\sigma(s_i|s_i) = 1$ and responsive voting strategies $\tau^M = \tau^*(M^\#)$:*

- $\tau^*(M^\#) = 1$ *if* $M^\# < \bar{S}$,
- $\tau^*(M^\#) = 0$ *if* $M^\# \geq \bar{S}$.

That is, under Unanimity, voting optimally aggregates information given truthful communication. The reason for this difference between Unanimity and Majority is that under Unanimity, 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.

Lemma 2 implies the following proposition:

Proposition 3 (Truthful Messaging under Unanimity)**.** *In any game $\hat{\Gamma}$ with d^v = Unanimity, an equilibrium exists with truthful messaging ($\sigma(s_i|s_i) = 1$) and committee-optimal decisions.*

Under Unanimity, all committee members must vote for an option for it to be selected. Therefore, even with the incentive to free-ride in the voting stage, given truthful communication it is a best-response for all agents to vote for the committee-optimal decision since no agent can deviate to a different voting strategy without preventing the committee from selecting the committee-optimal decision. Likewise, any deviation from truthful communication will either result in the same decision, or move the committee decision away from the optimal decision.

Propositions 2 and 3 allow us to make a comparison between Majority and Unanimity rules: relative to Unanimity, where an equilibrium exists with truthful communication and committee-optimal information aggregation for any level of the expressive payoff, if the expressive payoff is high all responsive equilibria under Majority lack both truthful communication and information aggregation. Next, we use the observations of Propositions 2 and 3 to inform our experimental design and test the qualitative predictions of the model.

3 Experiment

In this section, we explain how our experimental design may help us to detail the differences in communication and voting under Majority and Unanimity, and to explore the mechanisms behind the observed differences in aggregate behavior under the two decision rules.

3.1 Experimental Design

The experiment closely implements our model of voting with expressive payoffs and binary signals, using a 2×2 treatment design with "High" and "Low" expressive payoffs under Majority and Unanimity. 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 programmed 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 common-value payoff P_c, which committee members get after a committee decision that agrees 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 $K > K'$ 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 unanimity 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.[9] The experiment lasted between 75 and 105 minutes and subjects earned between 19 and 22 Euros

[9]This approach is known as the "pay all" method. Charness et al. (2016) review the evidence for and against "pay all" in contrast to the approach to "pay one" round and conclude that "in general, ... both are useful methods", suggesting that practical considerations such as potential bankruptcy (as in auctions, which would favor "pay one") should be considered when designing experiments. The practical considerations that let us to choose to "pay all" are the comparability with earlier experiments (such as Goeree and Yariv, 2011) and the potential introduction of background risk by "pay one", which would be a potential confound in our risky environment.

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
Unanimity-Low	Unanimity	40	10	45	4	50
Unanimity-High	Unanimity	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 4 in Appendix B.2 for a composite screenshot). After each game, the subjects received feedback on the state, payoffs, aggregate behavior and the aggregate signal profile. Under Majority, the timing of each round was as follows.

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

Under Unanimity, the timing of each round was identical to Majority aside from the voting stage. We did not allow for an infinite number of rounds of voting as in the theoretical model. Instead, subjects were given three chances to reach a unanimous decision, after which all subjects were assigned a default vote of R. This decision rule replicates the most important feature of Unanimity—that the final vote profile is uniform for any committee decision—and insures that a decision is reached in a reasonable time frame. This "default option" is the conservative choice in relation to our experimental hypotheses, as it ensures that the frequency of committee decisions equal to the non-expressive option under Unanimity is not driven by the default option. Additionally, analogous to the Unanimity rule, this game structure also admits an equilibrium in which all agents truthfully communicate their private messages and uniformly vote for the committee-optimal option.[10]

Unanimity The only difference to Majority is in the voting stage: Submit vote $v_i^1 \in \{R,B\}$. If vote is unanimous proceed to Outcome Stage. Otherwise, again submit a vote $v_i^2 \in \{R,B\}$. If vote is unanimous proceed to Outcome Stage. Otherwise, submit a final vote $v_i^3\{R,B\}$. If vote is not unanimous, all subjects are assigned the vote $v_i = R$. Proceed to Outcome stage.

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 and that this richer strategic space explains the higher degree of information aggregation we observe under Unanimity. We address this concern in a robustness section in the Appendix, and show that the higher degree of information aggregation observed under Unanimity is almost entirely driven by groups that reach a decision in the first round of voting, see Figure 7, which suggests that additional rounds of voting under Unanimity is not driving our results. Upon completion of the experiment, subjects left the laboratory and were paid individually in a separate room by an experimental assistant.

[10]This result follows as a corollary of Theorem 2—since the final profile of votes is homogenous in all decisions, any deviation from truthful communication and committee-optimal voting leaves agents weakly worse off. Additionally, since B is optimal only after three signals B for both high- and low-expressive payoffs, this decision rule admits an equilibrium where agents optimally aggregate information by babbling in the message stage and voting sincerely. As we show below, however, a relatively high number of subjects communicate truthfully under unanimity and respond to messages of other subjects in their voting behavior, and 66 percent of groups reach a unanimous decision in the first round of voting.

3.2 Theoretical Predictions and Experimental Hypotheses

Proposition 2 shows that for a wide range of expressive payoffs, truthful communication is ruled out in responsive equilibria under Majority voting. The expressive payoffs induced in the experiment, $10/40$ and $15/35$, are comparably small in the sense that they do not dominate the common-value payoffs from matching the committee decision to the state, but are large enough to satisfy $K \geq K'$ as defined in Proposition 2. That is, in both Majority treatments, expressive payoffs are large enough to suppress truthful communication in equilibrium, which yields the following hypothesis.

Hypothesis 1. *Messages are more likely to be truthful under Unanimity than under Majority after R signals but not after B signals.*

Proposition 2 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 1 shows 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 of the Majority rule.

As a result of the free-rider problem, Propositions 2 and 3 predict that Unanimity will outperform Majority in terms of optimal information aggregation. Additionally, based on the behavioral model 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 hypothesis.

Hypothesis 2. *Information is aggregated more efficiently under Unanimity when the committee-optimal decision is B but not when the committee-optimal decision is R.*

Conditional on resolving Hypotheses 1 and 2 as predicted, theoretically, the channel for the increased level of information aggregation under Unanimity could be twofold. On the one hand, committee members may anticipate and account for the more truthful messages under Unanimity. On the other hand, even if subjects hold the belief that their co-players' messages are truthful, Lemma 1 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.

Hypothesis 3. *Subjects anticipate and respond to more truthful messages under Unanimity.*

Second, Lemmas 1 and 2 predict that agents will best-respond to truthful reporting by free-riding, i.e. by misreporting R signals to induce 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.

Hypothesis 4. *A significant share of 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.

4 Analysis of the experiment

We address the four experimental hypotheses 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 Hypotheses 1 and 2

Expressive payoff	Low			High		
Decision Rule	Maj	Una	Diff	Maj	Una	Diff
Truthful message if *R* signal	90%	96%	0.057**	80%	94%	0.142**
Truthful message if *B* signal	84%	86%	0.018	87%	86%	-0.009
Committee-optimal decision *R*	80%	79%	-0.008	88%	86%	-0.019
Committee-optimal decision *B*	60%	83%	0.237**	61%	83%	0.217**

Note: The results on significance of differences between the treatments stem from linear probability models controlling for unobserved heterogeneity by multi-level random effects on sessions and subjects within sessions (in one-sided tests according to our hypotheses). Here, * denotes significance at 0.1 level, ** denotes significance at 0.05, and *** denotes signifance at 0.01.

4.1 Are messages more truthful under Unanimity?

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 evaluate Hypothesis 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 Unanimity 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 Unanimity and Majority are very small (at most two percentage points) and far from being significant, confirming the second part of Hypothesis 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, but is otherwise comparable). Focusing only on Unanimity, however, the relative frequency of truthful messages is comparable with Guarnaschelli et al. (2000).

Result 1. *As predicted (Hypothesis 1), messages are more truthful in Unanimity 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 expressive payoff treatment, from 10% to 21%. This finding is consistent with the quantitative increase in the incentive to strategi-

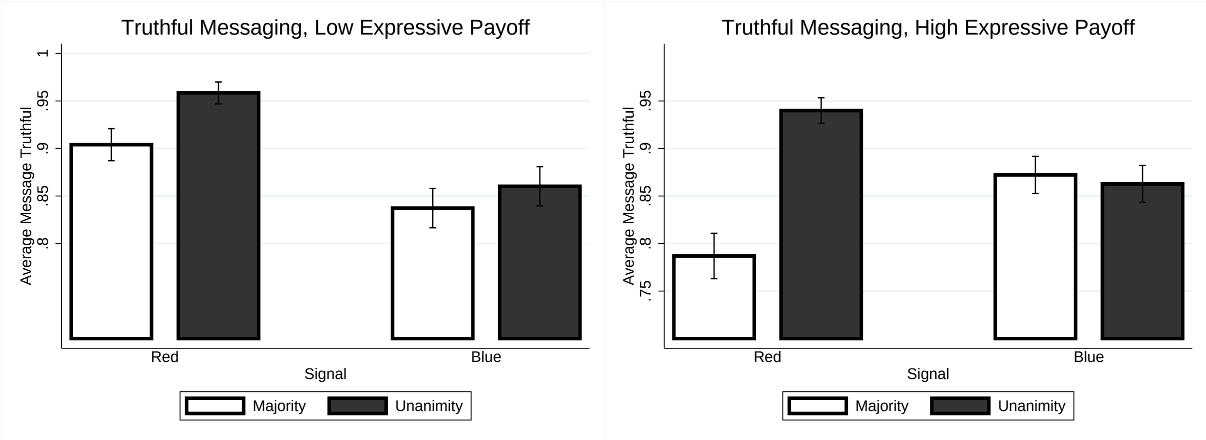


Figure 1: Average levels of truthful messaging ($m_i = s_i$) by signal and treatment.

cally 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 Unanimity. 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.2 in the Appendix further shows that the results are highly robust to experience.

4.2 Is information aggregated more efficiently under Unanimity?

While truthful communication is a necessary condition for the committee to behave optimally, the most pertinent comparison between Majority and Unanimity 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.[11] As detailed in Hypothesis 2, our prediction is that there is no difference between Unanimity and Majority if the expressive option, *R*, is committee optimal, but that under Unanimity the committee will be more likely to select the non-expressive option, *B*, when it is optimal.

[11] In the low-expressive payoff treatment, aggregate utility is maximized if the committee selects *B* given two signals of *B* and only two committee members vote for *B*.

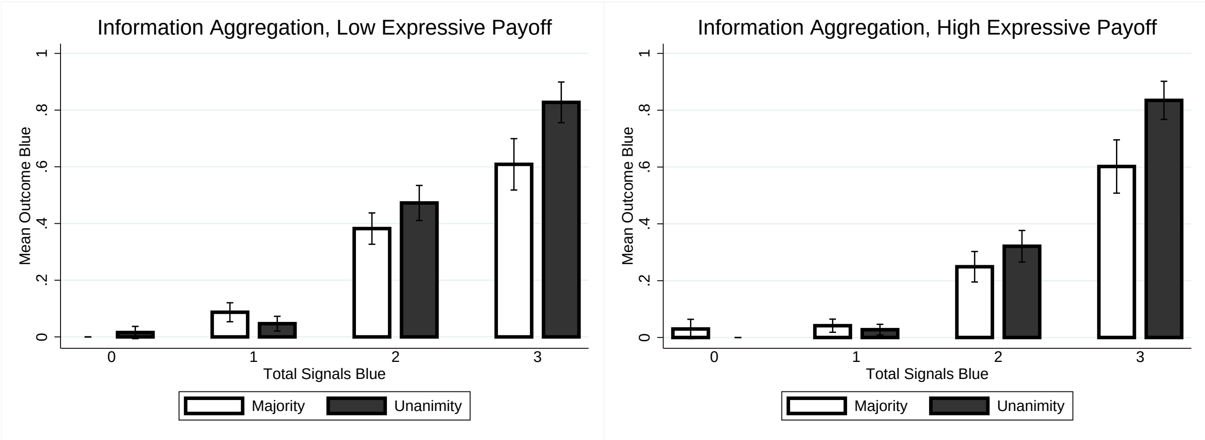


Figure 2: Average levels of outcome *B* as a function of the aggregate profile of *signals* (not messages) by treatment.

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 Unanimity 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 Unanimity.[12]

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

4.3 Do subjects anticipate more truthful messages under Unanimity?

Having established that Unanimity results in more truthful messaging and increased information aggregation (Hypotheses 1 and 2) we now turn to the more subtle questions regarding mechanism (Hypotheses 3 and 4). For a first pass at exploring the effect of messages under the

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

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 Unanimity, the impact of a message for *B* from a co-player has a similar 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 subjects anticipate the higher proportion of misreporting observed under Majority, which 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*	High
Own signal *B*	1.896*** (0.102)
Number of others' messages *B*	1.928 *** (0.0877)
Majority	1.453*** (0.238)
Majority × own signal *B*	−0.543*** (0.129)
Majority × others' messages *B*	−1.016*** (0.104)
Constant	−4.228*** (0.192)
N	4650

Note: We report standard errors in parentheses controlling for unobserved heterogeneity by multi-level random effects on sessions and subjects within sessions.* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

This direct comparison between the decision rules is suggestive of the predicted Hypothesis 3. 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 Unanimity, 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. We present this analysis in the Online Appendix, Section B, which confirms that subjects' beliefs react more strongly to co-players' messages under Unanimity:

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

4.4 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 first consider the average voting strategy of subjects in 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* (16 percent is also much lower than the analogous rate under Unanimity/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. Establishing this link is critical, as free-riding is not the only conceivable reason to misreport signals. Another reason is to attempt persuasion. For example, while our model assumes that agents are risk neutral, if subjects have heterogenous risk preferences, they may hold different preferences regarding the optimal option conditional on a given set of signals. Therefore, subjects who are, say, less risk averse than average may choose to misreport *R*-signals to increase the probability that the committee chooses *B* given two signals for *B*. The difference to free-riding is that *persuasive misreporting* is followed up by also voting for the misreported option (in case opponents sent two messages for *B*), while free-riders will vote *R* regardless of their opponents' messages.

Overall, it is easy to think of at least five classes of individual strategies: In addition to the three misreporting strategies 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"), which is our equilibrium prediction for Unanimity, 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.

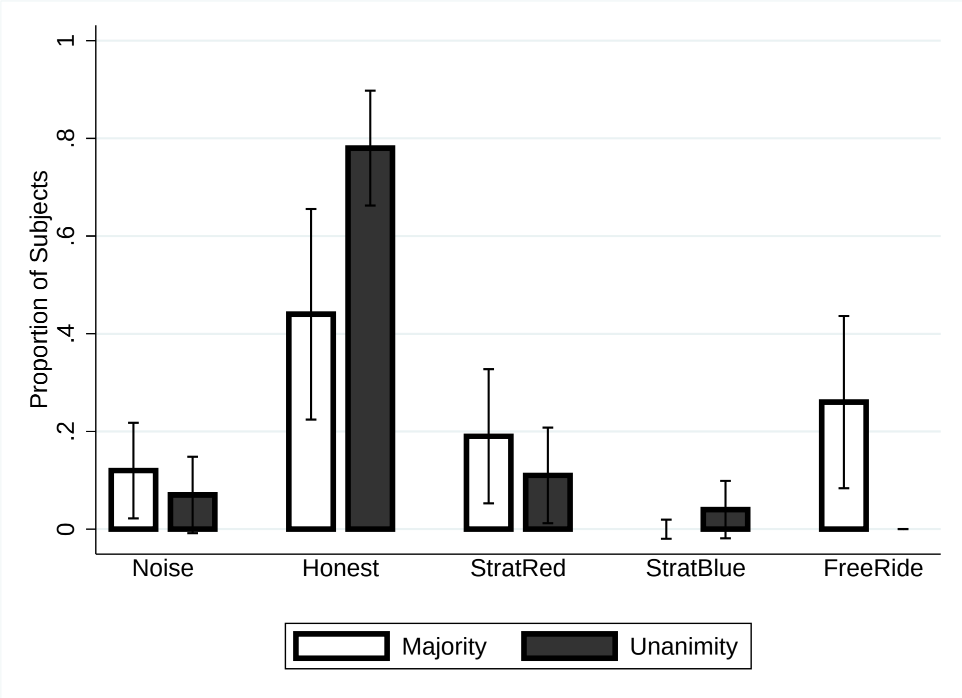


Figure 3: Classification of types by decision rule.

Table 5b in the Online Appendix presents the estimated strategy weights and strategy parameters, alongside the bootstrapped standard errors and statistical tests of our hypothesis. The main results are summarized in Figure 3, and allow us to conclusively conclude that:

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

To summarize, 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 Unanimity, 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.

5 Conclusion

In this paper, we explore strategic communication in committees that are subject to mixed motives, and the impact on the quality of committee's decisions. Using a theoretical model, we show that when the committee aggregates votes via a majority rule, there are no equilibria with truthful communication and responsive voting despite the fact that committee members have homogeneous preferences over the committee decision. In contrast, an equilibrium with truthful communication and committee-optimal decisions exists under unanimity rule for any expressive payoffs, as long as committee members have homogeneous preferences over which committee decision to take given the aggregate profile of signals. This finding also suggests a novel rationale for the use of a unanimity rule: in settings with expressive 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 unanimity 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

American Bar Association (December 02, 2013). How Courts Work. (Retrieved from https://www.americanbar.org/groups/public_education/resources/law_related_education _network/how_courts_work/jurydeliberate/).

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 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.

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

Callander, S. (2008). Majority rule when voters like to win. *Games and Economic Behavior*, 64:393–420.

Charness, G., Gneezy, U., and Halladay, B. (2016). Experimental methods: Pay one or pay all. *Journal of Economic Behavior & Organization*, 131:141–150.

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.

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

European Commission (2017). Proposal for a Council Regulation on the Establishment of the European Monetary Fund. Brussels.

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

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. (2018). How Transparency Kills Information Aggregation: Theory and Experiment. *American Economic Journal: Microeconomics*, 10(1):181–209.

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.

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.

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

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.

Maggi, G. and Morelli, M. (2006). Self-enforcing voting in international organizations. *American Economic Review*, 96(4):1137–1157.

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

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.

Sapir, A. and Schoenmaker, D. (2017). The Time is Right for a European Monetary Fund. *Bruegel Policy Brief*, 4:1–7.

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.

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.

Willman, D. (December 19, 2000). With study results kept secret, nation got another blood-pressure drug. *Los Angeles Times*, (Retrieved from http://www.pulitzer.org/winners/david-willman).

Appendix

Table of Contents

A Theoretical Appendix

A.1 Robustness

In order to conveniently discuss robustness and extensions, we continue to restrict our attention to the simple model with a binary signal space and an expressive payoff, K, for option R. In contrast to the analysis above, some of our results will be sensitive to the sign of K. Therefore, to facilitate the exposition of the results, we will explicitly consider both negative and positive expressive payoffs; i.e. $K \in (-1,0) \cup (0,1)$.

Additionally, we define $\sigma^* \equiv \sigma(s_i|s_i) = 1$ and $\tau^*(s_i, m_i, M^\#)$ as follows:

$$\tau^*(s_i, m_i, M^\#) = \begin{cases} 1, \text{ if } M^\# < \bar{S}, \\ 0, \text{ if } M^\# \geq \bar{S}. \end{cases}$$

where $\bar{S}$, defined in the previous subsection, is the minimum number of signals for B required for $X = B$ to be committee-optimal.

Non-homogenous expressive payoffs Intuitively, a robustness concern is posed by the possibility that committee members may not be subject to a homogenous expressive payoff. For example, elected representatives may face different expressive payoffs depending on their partisan affiliation and the composition of their individual districts. The following analysis shows that heterogeneity in the expressive payoff does not eliminate the collective action problem under Majority and, as long as K is not too large, then an equilibrium that results in committee-optimal outcomes exists under Unanimity.

Formally, we consider the following modification to our benchmark model: $N^+ \subset N$ receive a expressive payoff of K if $v_i = R$, and $N^- \subset N$ receive a expressive payoff of K if $v_i = B$. Members of these two groups use voting strategies $\tau^+(M^\#)$ and $\tau^-(M^\#)$, respectively. We assume without loss of generality that $|N^+| > |N^-|$, and assume that agents may condition their strategy on their expressive "type."

Proposition 4. *If d^v = Majority and committee members face heterogenous expressive payoffs, then for any symmetric equilibrium that exhibits truthful communication ($\sigma(s_i|s_i) = 1$) and responsive voting strategies $\tau^+(M^\#)$, $\tau^-(M^\#)$, there exists $S' \geq (N+1)/2$ such that:*

- $\tau^+(M^\#) = 1$, $\tau^-(M^\#) = 0$ *if* $M^\# < S'$,
- $\tau^+(M^\#) \in (0,1)$, $\tau^-(M^\#) = 0$ *if* $M^\# \geq S'$.

Proof: This proof follows the same logic as Lemma 1. If $M^\# < \bar{S}$, then each committee member maximizes their individual payoffs by voting according to their expressive payoff.

Next, for $M^\# \geq \bar{S}$, note that for a responsive equilibrium to exist, $\tau^+(M^\#)$ must be in $[0,1)$, for the majority expressive type. By the same logic as Lemma 1, $\tau^+(M^\#) = 0$ cannot be part of an equilibrium strategy, since in this case no agent is pivotal. Instead, there exists a $\tau^+(M^\#) \in (0,1)$ that constitutes a best response to truthful communication and $\tau^+(M^\#) = 0$. ■

The committee still faces the exact same free-rider problem under Majority with heterogenous expressive payoffs as with homogenous expressive payoffs (as illustrated by Lemma 1), with the one exception that the free-rider problem is limited to the set of agents in N^+—when the profile of signals indicates that B is optimal, then some agents in N^+ must vote B for the committee decision to equal B. Proposition 4 also illustrates that the free-riding problem under heterogenous expressive payoffs still result to an incentive to misreport for both types of agents. That is, since agents in N^+ play a mixed strategy when $M^\# \geq S'$ the committee is biased towards option R given truthful communication, which gives agents an incentive to misreport when they receive a signal of R.

Additionally, the optimality result under Unanimity is robust to heterogenous expressive payoffs, as long as K is small enough for the committee-optimal decision to be well-defined, in the sense that the optimal decision of a DM with access to the full profile of signals is independent of whether the DM receives a expressive payoff for voting R or for voting B. That is, take the following definition of $X^+(\mathbf{s})$ and $X^-(\mathbf{s})$.

Definition 4. *Take $X^+(\mathbf{s})$ ($X^-(\mathbf{s})$) to be the committee decision by a DM with preferences equivalent to that of agents in N^+ (N^-) who has access to the complete profile of signals, $\mathbf{s}$, and whose vote determines the committee decision, $X = v^{DM}$.*

The following result shows that an equilibrium exists with truthful communication and optimal committee decisions under Unanimity, even when expressive payoffs are heterogenous as long as K is small enough that $X^+(\mathbf{s}) = X^-(\mathbf{s})$ for all signal profiles.

Corollary 1. *In any game with heterogenous expressive payoffs, d^v = Unanimity and K such that $X^+(\mathbf{s}) = X^-(\mathbf{s})$, an equilibrium exists with truthful messaging ($\sigma(s_i|s_i) = 1$) and committee-optimal decisions.*

This corollary follows directly from Proposition 3: as long as the committee agrees on the optimal decision for each signal profile, then it is an equilibrium under Unanimity to truthfully communicate and vote for the optimal decision.

Non-cheap talk in the communication stage For two reasons, talk in the communication stage may not be cheap: lying aversion and expressive payoffs from messages. We begin with a comment on lying aversion. As we discuss in the introduction, many experimental papers have documented that individuals are hesitant to lie, even when lying is in their best interest. Note, however, that our prediction that Unanimity performs better than Majority when agents are subject to expressive payoffs does not rely on non-truthful communication.

Corollary 2 (Lying aversion)**.** *Assuming* $\sigma(s_i|s_i) = 1$ *(truthful communication), there are no symmetric equilibria with committee-optimal decisions under Majority.*

Corollary 2 follows directly from Lemma 1, which illustrates that agents face a coordination problem under Majority when the aggregate signal profile indicates that the non-expressive option is committee-optimal. This shows that the result that Unanimity outperforms Majority does not depend on the prediction of non-truthful communication (lying aversion does not impact the predictions of the model under Unanimity), and therefore is robust to lying aversion.

Related, in certain situations agents may also risk being exposed to expressive payoffs based on their communications to the committee. For example, if agents receive a negative reputation-based payoff for voting for B, then they might receive a similar payoff if it is revealed that they supported B in the communication stage. Next, we formally detail when communication payoffs prevent truthful communication under Unanimity.

To fix ideas, consider the following example: in addition to expressive payoffs, assume agents receive a negative payoff of $-K$ if they vote for B, *or if it is publicly revealed that they communicated support for B.* The probability that the committee's communication is publicly revealed is equal to δ. As in the baseline model, we normalize the expressive payoffs as a positive value for voting/communicating R. The expressive payoffs in the extended model are as follows:

$$u^2(X,\omega,v_i,m_i) = \begin{cases} \delta K, & \text{if } v_i = B, m_i = R \\ K, & \text{if } v_i = R, \\ 0, & \text{otherwise.} \end{cases}$$

The following proposition shows that Unanimity is robust to communication-based payoffs if the risk of communication becoming public is of positive but comparably small relevance next to one's final vote.

Proposition 5 (Truthful Messaging with Communication Payoffs)**.** *In any game* $\hat{\Gamma}$ *with* $K > 0$*,* $d^v =$ *Unanimity, and B is committee-optimal given* $S^\# = N$*, there exists a* δ' *such that if* $\delta < \delta'$*, an equilibrium of the game exists with truthful messaging and committee-optimal decisions.*

Proof: We prove the result by construction. By Corollary 3, $\tau^*(s_i,m_i,M^\#)$ is a best response to truthful communication.

Next, we show that σ^* is a best response to truthful communication given $\tau^*(s_i, m_i, M^\#)$ and δ small. If $s_i = R$, then $m_i = R$ is a best response by Corollary 3, since deviating to $m_i = B$ results in an additional payoff loss of δK. If $s_i = B$, then deviating to $m_i = R$ results in a payoff gain of δK. However, given $\tau^*(s_i, m_i, M^\#)$, deviating to $m_i = R$ results in $X = R$ when i's message is pivotal; i.e. $S^\# = \bar{S}$. Therefore, the relative expected payoff of deviating to $m_i = R$ is equal to:

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

Since this expression is nonpositive for $\delta \leq \delta' = (\Pr(S^\# = \bar{S}|s_i)/K)[\Pr(\omega = B|\bar{S}) - \Pr(\omega = R|\bar{S}) - K]$, truthful messaging is a best response to truthful communication and $\tau^*(s_i, m_i, M^\#)$ for $\delta \leq \delta'$.

Lastly, we show that there is no joint deviation in m_i, v_i that is a best response to $(\sigma^*, \tau^*(s_i, m_i, M^\#))$. Again, if $s_i = R$, then this follows from Corollary 3. If $s_i = B$ and $m_i = R$, then $\tau^*(s_i, m_i, M^\#)$ is a best response for $M^\# \neq \bar{S} - 1$. Moreover, the voting outcomes are equivalent to $(\sigma(s_i|s_i) = 1, \tau^*(s_i, m_i, M^\#))$ for all $S^\# \neq \bar{S} - 1$.

Therefore, the only possible deviation from $(\sigma(s_i|s_i) = 1, \tau^*(s_i, m_i, M^\#))$ that could be a best response is $m_i = R$, $\tau^*(s_i, m_i, M^\#))$ for $M^\# \neq \bar{S} - 1$ and $\tau(B, R, \bar{S} - 1)) = 0$. In this case, the committee would not reach an agreement when $M^\# \neq \bar{S} - 1$, and the relative expected payoff of deviating is:

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

which is strictly lower than zero for $\delta \leq \delta'$. ■

Note that even under Unanimity, agents' messages are not always pivotal for the committee decision. Therefore, agents have an incentive to report R given a signal of B, analogous to the incentive to vote R under Majority. In contrast to the voting stage, however, the probability of a message being pivotal is strictly positive even when other agents play pure (truthful) strategies. Therefore, truthful messaging can be supported as an equilibrium strategy even with communication-based payoffs as long as δ is not large.

Expressive payoffs in the coordination stage Briefly, let us also consider the related case where agents are exposed to expressive payoffs in the coordination stage, when they coordinate on the final voting decision (straw poll), and detail the robustness of Unanimity to this extension of the model. To make the contrast as stark as possible, we consider the case where agents receive a payoff of K for *any round in which* $v_i^t = R$. That is, given a unanimous decision is reached in round T, agents receive the following expressive payoffs:

$$u^2(X, \omega, \{v_i^t\}) = K \times \sum_{t=1}^{T} I_{v_i^t = R}.$$

The following proposition details the partial robustness of Unanimity rule to expressive payoffs for voting in the straw poll.

Proposition 6 (Truthful Messaging with Coordination Payoffs)**.**

1. *In any game $\hat{\Gamma}$ with $K < 0$ and $d^v =$ Unanimity, an equilibrium with truthful messaging and committee-optimal decisions exists where the committee reaches a decision in the*

first round of voting in the straw poll.

2. *In any game $\hat{\Gamma}$ with $K > 0$ and d^v = Unanimity, there is no stationary equilibrium with truthful messaging where the committee selects B with probability one when B is committee-optimal.*

Proof of Proposition 6: (1) If $K \in (-1, 0)$, then existence of an equilibrium with truthful communication and $\tau = \tau^*(s_i, m_i, M^\#)$ follows from the same logic as in the proof of Proposition 2: Any deviation that changes i's payoffs results in either (1) $X = B$ and $v_i = B$ with positive probability for some $S^\# < \bar{S}$, (2) $X = R$ with positive probability for some $S^\# \geq \bar{S}$, or (3) that the committee decision is delayed past the first round of voting for some $M^\#$, or some combination of all three. Since (1), (2), and (3) all imply a weak decease in i's expected payoffs, no deviation from $(\sigma^*, \tau^*(s_i, m_i, M^\#))$ is profitable.

(2) For $K \in (0, 1)$, take $M^{\#'} \geq \bar{S}$. If agents play $\tau^*(s_i, m_i, M^\#)$, then all agents vote B with probability one when $M^\# = M^{\#'}$. Therefore, if i deviates to $\tau(s_i, m_i, M^{\#'}) = 0$, no agreement is reached in any voting round when $M^\# = M^{\#'}$. Moreover, i receives a payoff of K in each round of voting, which is a profitable deviation. ■

Intuitively, when expressive payoffs are negative, then agents have an incentive to reach a decision in the first round of voting to avoid additional expressive payoffs that may accrue in any additional rounds of voting. When expressive payoffs are positive, however, then agents have an incentive to vote for as many rounds as possible, accruing positive expressive payoffs each time they vote for R. Therefore, while $v_i^t = R$ is a best response given $M^\# < (N+1)/2$ (deviating to $v_i = B$ is not individually optimal) when expressive payoffs are positive, if all agents vote for B with certainty for some $M^\#$, then agent i has a best response to deviate to $v_i^t = R$, since voting in the straw poll will proceed indefinitely and i will receive K in each round.

Status-Quo Unanimity rule In our baseline analysis, we consider a unanimity rule that requires that all committee members vote unanimously for an option to be chosen. This differs from a *status-quo* unanimity rule, where one option is designated as the status-quo ex ante, and is selected if one or more committee members vote for the status quo.

Definition 5 (Status-quo unanimity)**.** *Each player submits a vote v_i. The final decision X is B if all players vote B, and is R otherwise.*

In our analysis of the status-quo unanimity rule, we find that it is "robust" to positive (relative) payoffs for voting for the status quo, but faces similar problems as Majority when committee members face negative (relative) expressive payoffs.

Proposition 7 (Status-Quo Unanimity)**.**

1. *In any game $\hat{\Gamma}$ with $K > 0$ and d^v = Status-Quo Unanimity, there exists an equilibrium with truthful messaging and committee-optimal decisions.*
2. *In any game $\hat{\Gamma}$ with $K < 0$ and d^v = Status-Quo Unanimity, there is no symmetric equilibrium with truthful messaging and committee-optimal decisions.*

Proof of Proposition 7: For (1), by contradiction, assume there exists a profitable deviation from $(\sigma^*, \tau^*(s_i, m_i, M^\#))$, $(\sigma', \tau'(s_i, m_i, M^\#))$. For i to be better off, $(\sigma', \tau'(s_i, m_i, M^\#))$ must introduce a positive probability that either (i) $v_i = R$ for some $S^\# \geq \bar{S}$, or (ii) $X = B$ for some $S^\# < \bar{S}$ (or both)—any other changes in payoff-relevant outcomes would decrease i's expected utility. However, under status-quo unanimity, if $v_i = R$ then $X = R$, and if $X = B$ then $v_i = B$, which implies that both (i) and (ii) decrease i's expected utility.

(2) follows as a corollary to Proposition 1: Note that given $K < 0$, the free-riding condition is satisfied for any $S^\# < (N+1)/2$ other than $S^\# = 1$. Therefore, given truthful communication, there is no symmetric equilibrium in which $\Pr(X = R|S^\#) = 1$ for $S^\# \in \{0, [2, (N-1)/2]\}$ by the same argument as in the proof of Proposition 1. ■

When committee members face a positive payoff for voting for the status quo, then an equilibrium with truthful communication and committee-optimal voting exists under both non-status-quo and status-quo unanimity rules. For example, if committee members receive a positive expressive payoff for voting for R and option B is committee optimal, then all committee members must forgo the expressive payoff to choose the committee-optimal option. Thus, in the case of $K > 0$, a status-quo rule eliminates the collective-action problem that occurs under Majority when voting for B is committee optimal.

However, if committee members face a negative payoff for voting for the status quo, then a status-quo unanimity suffers from the same free-riding problem as majority. For example, if committee members receive a negative expressive payoff for voting for R and option R is committee optimal, then only one committee member must pay the negative expressive payoff under a status-quo unanimity rule to choose the committee-optimal option—that is, a status-quo unanimity rule functions as sub-majority rule for option R, requiring only a single vote for R to be selected. Therefore, analogous to a majority rule, given truthful communication there is no symmetric equilibrium in which the committee selects R with probability one whenever R is committee optimal.

Generalized Results Next, we show that our main results hold with a generalized state-space and a broad set of expressive payoffs. That is, we consider the following generalizations of our framework: Each agent receives a private signal regarding the state, s_i, drawn i.i.d. from a finite signal space S according to the probability distributions $p^\omega \in \Delta(S)$. Each signal is partially informative, $p^R(s) \neq p^B(s)$, but no signal perfectly reveals the state of the world in the sense that $p^\omega(s) > 0$ for all $s \in S$ and $\omega \in \Omega$. We define $\mathbf{S} = S^N$ and let $\mathbf{s} \in \mathbf{S}$ indicate a specific profile of signals. Prior to voting, agents engage in communication: agents simultaneously send messages, $m_i \in S$, that are publicly observable (agents are free to send any message, regardless of their private signal).

Consistent with our motivation of studying information aggregation with expressive payoffs, the committee members' payoffs are a function of the committee decision, the underlying state of the world, *and the agent's vote* v_i. Formally, agents' preferences are represented by a utility function, $u(X, \omega, v_i) : \{R, B, D\} \times \{R, B\}^2 \to \mathbb{R}$. For conceptual reasons, let us split up the utility function as above:

$$u(X, \omega, v_i) = u^1(X, \omega) + u^2(X, \omega, v_i),$$

where $u^1(X, \omega)$ represents the common-value payoff for matching the committee decision to the state of the world, and $u^2(X, \omega, v_i)$ represents the expressive payoffs. This formulation is fairly general, allowing for expressive payoffs that can vary based on the committee decision and the state of the world; for example, an FDA committee member may receive a negative reputation payoff only if a drug that is approved by the committee, is later shown to result in severe side-effects, *and* the committee member individually voted to approve the drug. We impose the restrictions that $u^1(\omega,\omega) > u^1(X,\omega)$ and $u(\omega,\omega,v_i) > u(X,\omega,v_i')$ for all v_i, v_i', ω and $X \neq \omega$, which implies that agents' payoffs from the committee choosing the option that matches the state is greater than any expressive payoff—these restrictions are not strictly necessary, but imply a payoff structure analogous to the standard models of information aggregation, where agents' primary motivation is to match the decision to the state.

We make one additional assumption regarding payoffs. Let $E_\omega[u(X',\omega,v_i = X)|\mathbf{s}]$ denote the expected utility, taken over the states of the world ω, when the committee decision is X', player i votes $v_i = X$ (X may equal X'), and signal profile is $\mathbf{s}$:

Assumption 1. *For all* $\mathbf{s}$*, there exists an decision* X' *such that* $E_\omega[u(X',\omega,v_i = X')|\mathbf{s}] > u(D,\omega,v_i) = 0$.

This assumption entails that, given knowledge of the signal profile $\mathbf{s}$, there exists a possible decision X' such that all committee members are better off if the committee reaches this decision rather than deliberating in perpetuity.

We denote the game by $\Gamma = \langle P_R, S, p^\omega, N, d^v, u\rangle$, and agents' strategies are pairs $(\sigma_i(m_i|s_i), \tau_i(s_i, m_i, \mathbf{m}))$. Lastly, we introduce a definition of what we refer to as the *Free-Riding property*. We call it the Free-Riding property since one of our main results shows that when expressive payoffs result in this property being satisfied, they cause an incentive to free-ride under Majority.

Definition 6 (Free-Riding). *A game* $\Gamma = \langle P_R, S, p^\omega, N, d^v, u\rangle$ *satisfies* Free-Riding *if there exists a vector of signals* $\mathbf{s}'$ *such that the following two conditions are satisfied:*

1. *Given a committee decision for* X^* *such that*

$$E_\omega[u(X^*,\omega,v_i = X^*)|\mathbf{s}'] > E_\omega[u(X',\omega,v_i = X')|\mathbf{s}'],$$

with $X' \neq X^*$*, agents individually prefer to vote for* X'*,*

$$E_\omega[u(X^*,\omega,v_i = X')|\mathbf{s}'] > E_\omega[u(X^*,\omega,v_i = X^*)|\mathbf{s}'].$$

2. *Take* N^s *to be the set of agents with signal* s *given* $\mathbf{s}'$*; there is no signal set* $\hat{S} \subseteq S$ *such that* $|\cup_{s\in\hat{S}} N^s| = (N+1)/2$*, i.e. such that the agents with signals in* $\hat{S}$ *constitute a minimal winning coalition.*

The first condition of the Free-Riding property requires that expressive payoffs are non-trivial in the sense that they may induce a conflict between collective and individual interest for at least one signal profile. Absent expressive payoffs ($u^2 = 0$), Γ does not satisfy this condition since agents' utilities are independent of their individual votes: $u(X,\omega,v_i \neq X) = u(X,\omega,v_i = X)$ for all ω. In contrast, a simple expressive payoff for voting for R satisfies condition 1 straightforwardly since, holding the committee decision constant, agents receive a higher payoff from individually voting for R regardless of $\mathbf{s}$.

The second condition of the Free-Riding property implies that agents cannot always coordinate on who may free-ride based solely on their signals $\mathbf{s}'$. As a result, there can be no symmetric pure strategy equilibrium where the winning option always receives exactly $(N+1)/2$ votes under Majority. For example, a simple expressive payoff for voting for R satisfies both conditions of the Free-Riding property since for any homogenous signal vector ($\mathbf{s}'$ such that $s'_i = s'_j$ for all i, j) where option B maximizes the expected value of u^1, each agent prefers that the committee selects B, but individually prefers to vote R.

Main Results: Our first main result illustrates the free-riding problem that can occur under Majority when committee members are subject to expressive payoffs. It establishes that the Free-Riding property is indeed a sufficient condition for the non-existence of a symmetric equilibrium with truthful communication and committee-optimal decisions.

Theorem 1 (Non-existence under Majority)**.** *For any game* Γ *with* d^v = *Majority satisfying* Free-Riding*, there does not exist a symmetric equilibrium with truthful communication* ($\sigma(s_i|s_i) = 1$) *and committee-optimal decisions.*

In a game that satisfies *Free-Riding*, expressive payoffs introduce a coordination problem due to the fact that, under majority rule, it is possible for the committee to select a given option even when a minority of agents vote for the other option. Therefore, under Majority it will not be a best response for all agents to vote for the committee-optimal decision given a signal profile of $\mathbf{s}'$, since a single agent can deviate to voting for the other option and receive a higher expected utility. This implies that, given truthful communication, it is not a best-response for agents to play a symmetric voting strategy that selects the committee-optimal decision for all signal profiles.

Proof of Theorem 1: We prove the result by contradiction. Assume an equilibrium, (σ^*, τ^*) exist with truthful communication and committee-optimal outcomes. Given truthful communication, $\mathbf{m} = \mathbf{s}$. Therefore, for each $\mathbf{m}$, the associate set of $\tau^*(s_i, m_i, \mathbf{m})$ must imply that the committee selects option R (B) with probability one if R (B) is committee-optimal given $\mathbf{m}$.

Take $\mathbf{s} = \mathbf{s}'$, and take R to be the committee optimal outcome given $\mathbf{s}'$ (the same argument holds when B is committee optimal). In this case, $\Pr(X = R|\tau^*(s_i, m_i, \mathbf{s}')) = 1$, which implies that a majority of agents must vote $v_i = R$ with probability one.

Given condition (2) of *Free-Riding*, however, there is no partition of $\mathbf{s}'$, $\mathbf{s}_1, \mathbf{s}_2$, such that all agents with the same signal are in the same subset of $\mathbf{s}'$, and $\sum^i I_{s'_i \in \mathbf{s}_1} = (N+1)/2$. Therefore, since (σ^*, τ^*) is a symmetric pure strategy that selects $X = R$ with probability one when $\mathbf{s} = \mathbf{s}'$, it must be the case that $\tau^*(s_i, m_i, \mathbf{s}')$ specifies that more than $(N+1)/2$ agents play $v_i = R$. However, given $X = R$, it is individually optimal to vote for B for $\mathbf{s} = \mathbf{s}'$. Therefore, it is a best reply for agents with $\tau(s_i, m_i, \mathbf{s}') = 1$ to deviate to $v_i = B$, given that no individual agent's vote is pivotal for the committee decision. This implies that (σ^*, τ^*) cannot be an equilibrium. ■

Theorem 1 starkly contrasts with our main result for Unanimity.

Theorem 2 (Existence under Unanimity)**.** *For any game* Γ *with* d^v = *Unanimity, there exists a symmetric equilibrium with truthful communication and committee-optimal decisions.*

Proof of Theorem 2: First, note that Unanimity limits the outcome set to points with $v_i = X$, or perpetual delay. Therefore, given utility functions of the form $u : \{X, \omega, v_i\} \to \mathbb{R}$, 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 $\sigma(s_i|s_i) = 1$, and $\tau(s_i, m_i, \mathbf{m}) = 1$ for $\mathbf{m} \in \mathbf{S}^R$ and $\tau(s_i, m_i, \mathbf{m}) = 0$ for $\mathbf{m} \in \mathbf{S}^B$ that results in a strictly positive increase in i's expected payoffs, where $\mathbf{S}^R$ and $\mathbf{S}^B$ are defined as in the main text as the subset of signal profiles where agents prefer $X = R$ and $X = B$ respectively given $v_i = X$.

Given that Unanimity limits the outcome set to points such that $v_i = X$ (or perpetual delay), 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. 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 $\mathbf{s} \in \mathbf{S}^B$, (3) that no decision is reached for some $\mathbf{s}$ ($X = D$), or some combination of all three. However, by the definition of $\mathbf{S}^R$, $\mathbf{S}^B$ and Assumption 1, all payoff-relevant changes (1), (2) and (3) imply that i's expected payoffs decrease or remain constant, which is a contradiction. ■

A.2 Proofs for Section 2.2

Before proceeding with the proofs for Section 2.2, we introduce some additional notation. First, let $piv(x, M)$ denote the probability that i's vote is pivotal under Majority when $M - 1 \in \{(N-1)/2, (N+1)/2, ..., N-1\}$ agents vote R with probability x, and all other agents vote R with probability one:

$$piv(x,M) = \frac{(M-1)!}{\left(\frac{N-1}{2}\right)!\left(M - \frac{N+1}{2}\right)!} x^{(N-1)/2} (1-x)^{M-(N+1)/2},$$

Next, let $\bar{S}$ denote the minimum number of total signals for B ($S^{\#}$) required for B to be committee optimal. That is:

$$\bar{S} = \begin{cases} N+1, \text{ if } \Pr(\omega = B|S^{\#}) - \Pr(\omega = R|S^{\#}) < K \;\; \forall \;\; S^{\#}, \\ \min\{S^{\#} \mid \Pr(\omega = B|S^{\#}) - \Pr(\omega = R|S^{\#}) \geq K\}, \text{ else.} \end{cases}$$

Committee-optimal information aggregation prescribes that the committee play strategies such that they select option R when the actual $S^{\#}$ is smaller than $\bar{S}$, and option B when $S^{\#}$ is greater or equal to $\bar{S}$.

Proof of Proposition 1: First, given that $\tau_j(,,,) = 1$ for $j \neq i$, the committee outcome is independent of v_i, which implies that $\Pr(X = \omega|v_i = R, M^{\#}, m_i, s_i) + K > \Pr(X = \omega|v_i = B, M^{\#}, m_i, s_i)$. Therefore, $\tau_i(,,,) = 1$ is a best response.

Second, given $\tau_i(,,,) = 1$, the committee outcome is independent of m_i, which implies that for any messaging behavior, a corresponding equilibrium exists with $\tau_i(,,,) = 1$. ■

Proof of Lemma 1: Define S' as follows:

$$S' = \begin{cases} N+1, \text{ if } piv(0.5,N)[\Pr(\omega = B|S^{\#}) - \Pr(\omega = R|S^{\#})] < K \ \forall \ S^{\#}, \\ \min\{S^{\#} \,|\, piv(0.5,N)[\Pr(\omega = B|S^{\#}) - \Pr(\omega = R|S^{\#})] \geq K\}, \text{ else.} \end{cases}$$

For $M^{\#} < S'$, the following expression holds for all $x \in [0,1]$ since $piv(x,N)$ is maximized at $x = 0.5$:

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

This implies that for $M^{\#} < S'$, $piv(\tau^M(M^{\#}),N)\Pr(X = \omega|v_i = R, M^{\#}, m_i, s_i) + K > piv(\tau^M(M^{\#}),N)\Pr(X = \omega|v_i = B, M^{\#}, m_i, s_i)$ for any $\tau^M(M^{\#})$ and that there is no truthful equilibrium in which $\tau^M(M^{\#}) < 1$ is a best response for $M^{\#} < S'$.

For $M^{\#} \geq S'$, note that $piv(x,N)$ is equal to zero for $x = 0,1$, and $piv(x,N)$ is strictly increasing over the domain $[0,0.5)$ and strictly decreasing over $(0.5,1]$. Therefore, there exists a unique $\tau^M(M^{\#}) \in (0,0.5]$ such that the best-response condition holds, and this $\tau^M(M^{\#})$, denoted as $\tau^*(M^{\#})$, is implicitly characterized by the following equation:

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

Additionally, by the symmetry of $piv(x,N)$ about 0.5, $piv(1-\tau^*(M^{\#}),N)[\Pr(\omega = B|M^{\#}) - \Pr(\omega = R|M^{\#})] = K$, and $1-\tau^*(M^{\#})$ is the unique equilibrium in $[0.5,1)$. ■

Lemma 1 characterizes the form of equilibria with truthful communication and responsive voting given the assumption that voting behavior does not conditional on the individual signal and message; i.e. $\tau(s_i, m_i, M^{\#}) = \tau^M(M^{\#})$. Below, we extend Lemma 1 to include equilibria with truthful communication and message-contingent voting. First, we define $S''(M^{\#})$ as follows:

$$S''(M^{\#}) = piv\left(\frac{N-1}{2(M-1)}, M^{\#}\right)\left[\Pr(\omega = B|S^{\#} = M^{\#}) - \Pr(\omega = R|S^{\#} = M^{\#})\right] - K.$$

Lemma 1' (Voting Stage: Majority, Message-Contingent)**.** *Given $d^v = Majority$, equilibria with truthful communication, $\sigma(s_i|s_i) = 1$, and responsive voting either take the form outlined in Lemma 1 (uniform voting), or the following form (message-contingent voting):*

- $\tau^*(R,R,M^{\#}) = 1$.
- $\tau^*(B,B,(N+1)/2) = 0$ *if* $\bar{S} = (N+1)/2$.
- $\tau^*(B,B,M^{\#}) \in (0,1)$ *if* $M^{\#} \geq S''(M^{\#})$.
- $\tau^*(B,B,M^{\#}) = 1$ *else.*

Proof of Lemma 1': We begin by showing that $\tau^*(B,B,M^{\#})$ constitutes a best response given $\tau(R,R,M^{\#}) = 1$. First, if $\bar{S} = (N+1)/2$ and $M^{\#} = (N+1)/2$, then it is a best response for i to set $v_i = B$ if i is pivotal with probability one. Therefore, $\tau^*(B,B,(N+1)/2) = 0$ is a best response since $\tau(R,R,M^{\#}) = 1$, ensuring that i with $s_i = B$ is pivotal when $M^{\#} = (N+1)/2$.

Next, note that $piv(x,M)$ is strictly increasing over the domain $[0,(N-1)/(2(M-1)))$ and strictly decreasing over $((N-1)/(2(M-1)),1]$. Therefore, given $M^{\#} \geq S''(M^{\#})$ there exists

$\tau^*(B,B,M^\#) \in (0,(N-1)/(2(M-1)))$ that is a best response by the same argument as in the proof of Lemma 1.

Likewise, if $M^\# < S^{''}(M^\#)$ and $M^\# \neq (N+1)/2$, then $\tau(B,B,M^\#) = 1$ is a best response to any $\tau^*(B,B,M^\#) \in [0,1]$.

Next, we show that $\tau(R,R,M^\#) = 1$ is a best response given $\tau^*(B,B,M^\#)$. If $M^\# < S^{''}(M^\#)$ or $M^\# = (N+1)/2$, then i with $s_i = R$ is never pivotal, which implies that $v_i = R$ is a best response. If $M^\# \geq S^{''}(M^\#)$, then $\tau(R,R,M^\#) = 1$ is a best response if:

$$piv\big((1-\tau^*(B,B,M^\#)),M^\#+1\big) < piv\big((1-\tau^*(B,B,M^\#)),M^\#\big)$$

Simplifying this expression gives:

$$\frac{N-1}{2} < M^\#(1-\tau^*(B,B,M^\#)),$$

which holds for $\tau^*(B,B,M^\#) \in (0,(N-1)/(2(M-1)))$ and $M^\# \geq (N+3)/2$.

Additionally, note that for $\tau(R,R,M^\#) \in (0,1)$ and $\tau(B,B,M^\#) \in (0,1)$ to constitute a best response for a given $M^\#$, the probability of being pivotal for i with $s_i = R$ must equal the probability of being pivotal for i with $s_i = B$. This implies that any equilibrium with truthful communication and responsive voting for both $s_i = R$ and $s_i = B$ take the form $\tau(R,R,M^\#) = \tau(B,B,M^\#)$ (i.e. $\tau(s_i,m_i,M^\#) = \tau^M(M^\#)$).

Lastly, $\tau(R,R,M^\#) \leq 1$ and $\tau(B,B,M^\#) = 1$ and $\tau(R,R,M^\#) < 1$ for some $M^\#$ cannot constitute a best response, since given $\tau(R,R,M^\#) < 1$ and $\tau(B,B,M^\#) = 1$, the probability that i's vote is pivotal is only greater than zero for $M^\# < (N+1)/2$. However, when $M^\# < (N+1)/2$, then $\tau(R,R,M^\#) = 1$ maximizes payoffs independent of other's voting behavior. Therefore, all equilibria with truthful communication and responsive voting take the form $\tau(R,R,M^\#) = \tau(B,B,M^\#)$ (Lemma 1), or $\tau(R,R,M^\#) = 1$ and $\tau(B,B,M^\#) \leq 1$ (Lemma 1'). ■

Proof of Proposition 2: We provide a general proof of Proposition 2 by analyzing three separate cases.

CASE 1: $\tau(s_i,m_i,M^\#) = \tau^M(M^\#)$: This case corresponds to the form of equilibria we focus on in the main body of the text.

We prove the result by contradiction. Assume an equilibrium exists with truthful messaging in the deliberation stage, and that all agents play strategy $\tau^M(M^\#) = \tau^*(M^\#)$ in the voting stage. Note that the formulation of the Lemma implies that given K, $\tau^*(M^\#) > 0$ for $M^\# = N$ (that is, $S' \leq N$, otherwise the only equilibrium is the non-responsive equilibrium).

Let $\bar{K}$ denote the difference in the posterior probabilities of both states conditional on $(N+1)/2$ signals for B.

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

That is, $\bar{K}$ is the relative expected utility that i receives when the committee selects option B, given $(N+1)/2$ signals for B. Additionally, take $K' = piv(0.5,N)\bar{K}$.

The proof stems from the following two observations: (1) since $K > K'$ implies that $S' > (N+1)/2$, when $M^\# = (N+1)/2$, the unique best response is for all agents to vote for R ($\tau^*(M^\#) = 1$); (2) by Equation 1 in the proof of Lemma 1, $\tau^*(M^\#)$ is decreasing 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), $\tau^*(M^\#) = 1$ for $M^\# \leq (N+1)/2$, and therefore $\Pr(X = B | S^\# \leq (N-1)/2, m_{i'} = B) = 0$, which implies that despite the deviation, B will never be chosen when R receives more signals. Second, by (2), $\Pr(X = B | S^\# \geq (N+1)/2, m_{i'} = B) > \Pr(X = B | S^\# \geq (N+1)/2, m_{i'} = R)$ for $S^\# > S' - 1$; i.e. the probability that B is chosen when B receives the majority of signals is higher given i''s deviation.

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 $M^\#$ $(\sigma(R) = \sigma(B) = 0, \tau(M^\#) = 1)$).

This shows that for $K > K' = piv(0.5, N)\bar{K}$, there is no equilibrium with truthful communication where agents play $\tau^*(M^\#)$ in the voting stage.

CASE 2: $\tau(s_i, m_i, M^\#) = \tau^M(M^\#), (1 - \tau^M(M^\#))$), $\tau(M^\#) = 1$: Note that it is possible that equilibria with truthful communication feature voting behavior that, e.g., alternates between $\tau^*(M^\#)$, $(1 - \tau^*(M^\#))$ and $\tau(M^\#) = 1$ (non-responsive) for $M^\# \in \{(N+3)/2, (N+5)/2, ..., N\}$. In this case, $\Pr(X = B | M^\#, \tau(s_i, m_i, M^\#))$ is not necessarily weakly increasing in $M^\#$, which implies that the proof provided for Case 1 does not apply, since i' with $s_{i'} = R$ deviating to $m_{i'} = B$ could lower the probability that B is selected for some $M^\#$.

However, for large enough K, it will still be a best response for i with a signal of R to deviate to $m_i = B$. Take S' defined as above, i.e.:

$$S' = \begin{cases} N+1, \text{ if } piv(0.5, N)[\Pr(\omega = B | S^\#) - \Pr(\omega = R | S^\#)] < K \;\; \forall \;\; S^\#, \\ \min\{S^\# \,|\, piv(0.5, N)[\Pr(\omega = B | S^\#) - \Pr(\omega = R | S^\#)] \geq K\}, \text{ else.} \end{cases}$$

To see that a lower bound on K exists such that there are no equilibria with truthful communication and responsive voting for K larger than this bound, take K large enough so that $S' = N$. By Lemma 1, the only possible responsive equilibrium that features truthful communication and uniform voting has $\tau^*(M^\#) < 1$ iff $M^\# = N$. Therefore, given truthful communication, there is a positive probability that the committee selects B only if $S^\# = N$. However, in this case, it is not a best response for i' with $s_{i'} = R$ to set $m_{i'} = R$, since a unilateral deviation to $m_{i'} = B$ results in a positive probability that the committee selects B when $S^\# = N - 1$. Since this argument holds regardless of whether agents play $\tau^*(M^\#)$ or $(1 - \tau^*(M^\#))$ (given $S' = N$, agents must vote responsively for $M^\# = N$ for the equilibrium to be responsive) in the voting stage, this shows that given K large enough so that $S' = N$, no equilibria exist with truthful communication and responsive voting.

CASE 3: MESSAGE-CONTINGENT VOTING ($\tau^*(R, R, M^\#) = 1, \tau^*(B, B, M^\#)$): Next we consider the case of message-contingent voting outlined in Lemma 1'. First, note that similar to Case 2, there may exist two values of τ that constitute a best response for $M^{\#} \geq S''(M^{\#})$, $\tau'(B,B,M^{\#}) \in (0,(N-1)/(2(M-1)))$ and $\tau''(B,B,M^{\#}) \in ((N-1)/(2(M-1)),1)$. Also, by the same argument as in Case 1, if $K > [\Pr(\omega = B|S^{\#}) - \Pr(\omega = R|S^{\#})]$ for $S^{\#} = (N+1)/2$, then there are no equilibria with message-contingent voting and $\tau^*(B,B,M^{\#}) = \tau'(B,B,M^{\#})$ for $M^{\#} \geq S''(M^{\#})$, since agents with $s_i = R$ would profit from a deviation to $m_i = B$.

Additionally, equilibria where agents play $\tau'(B,B,M^{\#})$ for some $M^{\#} \geq S''(M^{\#})$, and $\tau''(B,B,M^{\#})$ for other $M^{\#} \geq S''(M^{\#})$ are ruled out for K high enough by the proof of Case 2, since strategies in the two cases are equivalent when $S' = N$.

NOTE: $N = 3$: Lastly, we note that when $N = 3$ (the case we consider in our experiment), if $K > \Pr(\omega = B|S^{\#}) - \Pr(\omega = R|S^{\#})$ for $S^{\#} = 2$, then by Cases 1-3, no symmetric equilibrium exists with truthful communication and responsive voting. ■

Proofs of Lemma 2 and Proposition 3: Both results follow from our generalized result, Theorem 2, which we proved above.

B Empirical Appendix

In line with Hypothesis 3, our hypothesis is that subjects' beliefs are more responsive to co-players' messages under Unanimity. We test this hypothesis by estimating a simple structural equation model of the decision making process.[13] 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 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[14]

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

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 $\beta_{2|1}$, $\beta_{2|2}$,

[13] 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.

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

$\beta_{2|3}$ 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).

In our analysis, we allow for two forms of behavioral biases: (1) the belief weights α_1,α_2 are merely proportional (jointly) to their rational expectation counterparts, which captures biases due to overshooting ($\alpha_i>\hat{\alpha}_i$) and conservative belief formation ($\alpha_i<\hat{\alpha}_i$), and (2) the vote weights $\beta_{2|\cdot}$ are merely proportional to their empirical counterparts, which allows for biases due to overestimating ($\beta_i>\hat{\beta}_i$) and underestimating ($\beta_i<\hat{\beta}_i$) the predictability of the co-players' votes. In the extreme case, $\beta_{2|\cdot}=0$, subjects believe their co-players are perfectly unpredictable (usually dubbed level-1, see 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, our Hypothesis 3 that beliefs react more strongly to the co-players' messages under Unanimity, since messages are more truthful in this case, amounts to

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

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 hypothesis (resampling at the subject level to account for the panel character of the data, stratifying to acknowledge the treatment

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 R signal and two B messages from others				
$\Pr_{\text{Maj}}(\omega = R \mid s = R, m_2 = m_3 = B)$				
$\Pr_{\text{Una}}(\omega = R \mid s = R, m_2 = m_3 = B)$				
Beliefs after B signal and zero B messages from others				
$\Pr_{\text{Maj}}(\omega = R \mid s = B, m_2 = m_3 = R)$				
$\Pr_{\text{Una}}(\omega = R \mid s = B, m_2 = m_3 = R)$				
Log-Likelihood				
Robustness checks				
Rational state beliefs			✓	
Rational voting beliefs				✓
Second halves of sessions				
Beliefs after R signal and two B messages from others				
$\Pr_{\text{Maj}}(\omega = R \mid s = R, m_2 = m_3 = B)$				
$\Pr_{\text{Una}}(\omega = R \mid s = R, m_2 = m_3 = B)$				
Beliefs after B signal and zero B messages from others				
$\Pr_{\text{Maj}}(\omega = R \mid s = B, m_2 = m_3 = R)$				
$\Pr_{\text{Una}}(\omega = R \mid s = B, m_2 = m_3 = R)$				
Log-Likelihood				
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 unanimity treatments is robust nonetheless. As before, significance of differences between majority and unanimity estimates is indicated by plus and minus signs ($^{++/--}$ at $p < .05$ and $^{+/-}$ at $p < .1$ using bootstrapped p-values), next to the unanimity treatment estimates. All standard errors are bootstrapped.

structure). Parameters are estimated by maximum likelihood[15] 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 Unanimity, respectively. This shows that messages should be given weight—and more so under Unanimity treatments than under Majority, as the empirical probabilities deviate relatively more from 50-50 under Unanimity. 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 unanimity 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 and account for the higher probability of truthful messages under Unanimity.

B.1 Finite Mixture Modeling to Estimate Strategy Weights

The voting strategies we assign to the honest/naive type follow from the theoretical predictions of Lemma 1 and in all cases, we allow for noise. 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*.[16] 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.

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,[17] but is otherwise comparable to cluster analyses.[18]

[15] 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.

[16] 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.

[17] 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.

[18] 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.

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 $\rho = (\rho_1, \rho_2, \dots)$ 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 assignment 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. 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 parameterization. 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.

Note that our results are robust to pooling the Majority treatments and Unanimity 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 7 (see Appendix B.2), 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.

B.2 Additional Material and Robustness of the Experimental Results

Additional graphs and tables Figure 4 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.

Table 5: Do subjects play more honest strategies in unanimity 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 R given the signal $s \in \{R,B\}$. $\pi(s,m,M)$ is the probability of voting R 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					ICL-BIC
	Noise	Honest	StratRed	StratBlue	FreeRide	ε	π_{Lie}	π_{High}	π_{Med}	π_{Low}	
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 unanimity treatments and more free-riding 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 unanimity 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 unanimity 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.

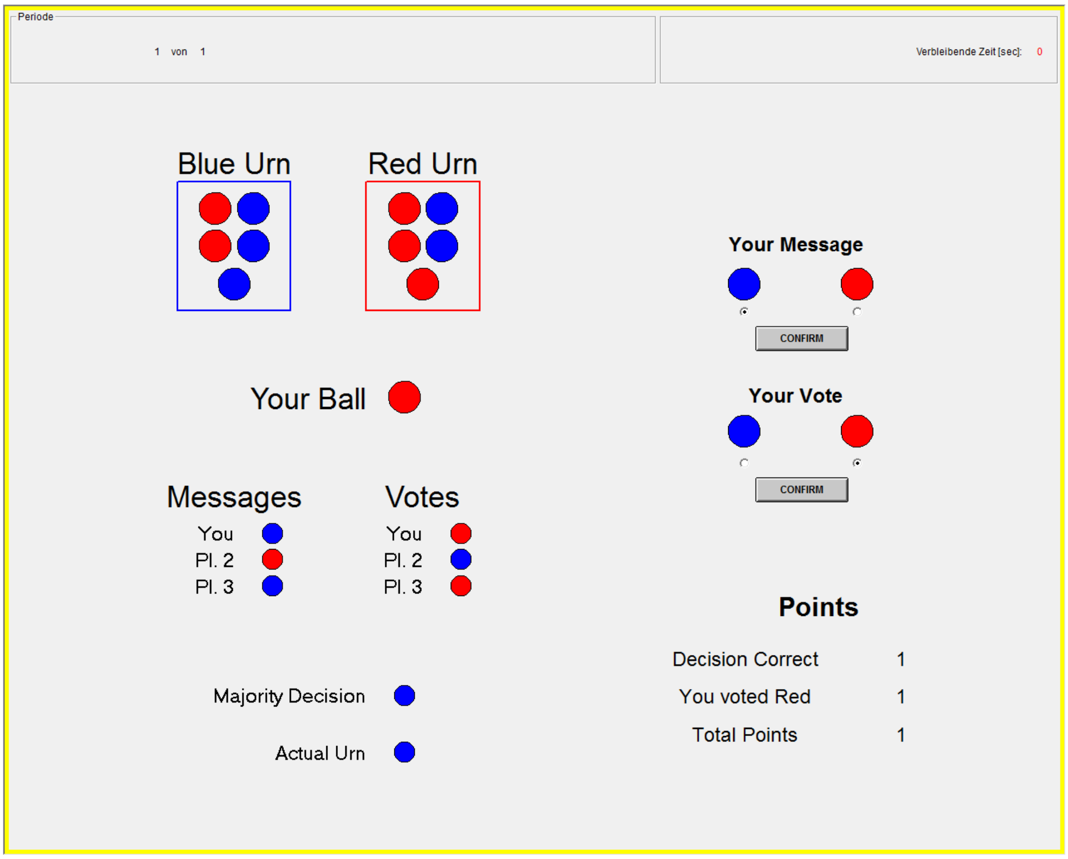
Learning In this subsection we show that the qualitative results presented in above are robust to excluding the first 25 rounds of the experiment (learning). Figures 5 and 6 replicate the Figures referred to when discussing Hypotheses1 and 2 for the restricted data set comprising only the second halves of all sessions. The patterns are virtually indistinguishable. The remaining statistical results relied upon in the discussion of Hypotheses 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 Unanimity As we discuss in the main text, there is an asymmetry in our operationalization of Majority and Unanimity. In particular, to replicate a unanimity decision rule that requires a unanimous decision and allows multiple rounds of voting, we allowed for up to three rounds of voting in the Unanimity treatment. Here we explore whether the multiple rounds of voting had a direct impact on the level of information aggregation—in particular, whether multiple rounds of voting contribute to the higher degree of information aggregation observed under Unanimity. To address this question, we compare the level of information aggregation between groups that reached a unanimous decision in the first round of voting and groups that required multiple rounds of voting to reach a unanimous decision or who did not reach a unanimous decision after three rounds of voting.

If the multiple rounds of voting were instrumental in driving the higher degree of information aggregation under Unanimity, then we would expect to see that groups that voted for multiple rounds did a better job at aggregating information relative to groups that reached a unanimous decision in the first round. However, as seen in Figure 7, this is not the case: the higher degree of information aggregation observed under Unanimity is entirely driven by groups that reached a unanimous decision in the first round of voting.

Finite mixture model: robustness checks Table 7 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.

Figure 4: Translated screenshot



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 Unanimity if decision unanimous or the third vote was taken) or disappear (in Unanimity 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 artificial. The information remains on the screen for 10 seconds, after which a new game starts.

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.

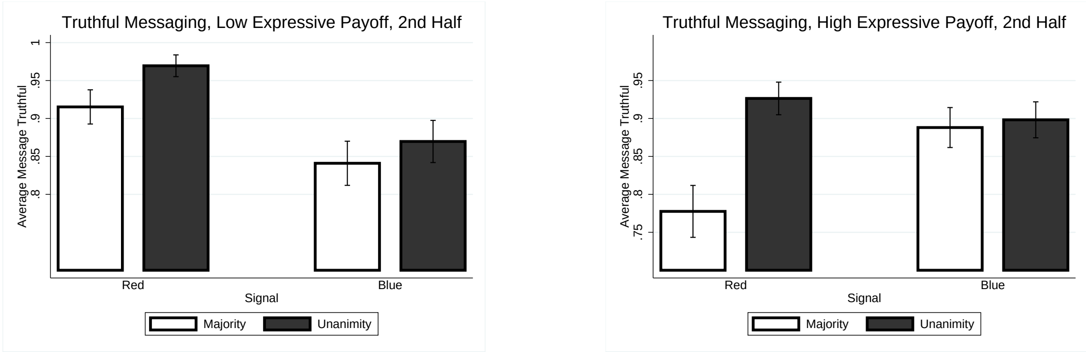


Figure 5: Truthful reporting for rounds $26-50$.

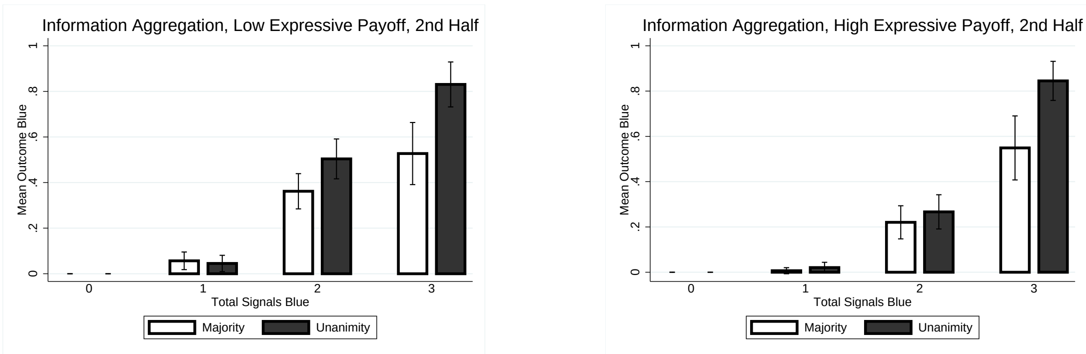


Figure 6: Information aggregation for rounds $26-50$.

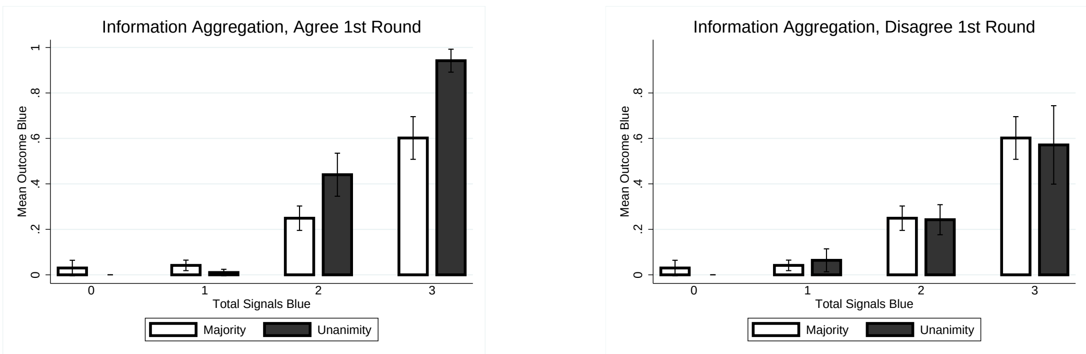


Figure 7: Information aggregation by signal profile, decomposed by whether the group reached a decision in the first round (left figure), or voted for multiple rounds (right graph).

Table 7: 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								