Honest Cheap Talk and Overweighting of Information in Committees*

Yves Breitmoser† Justin Valasek‡

January 6, 2026

Abstract

We study information aggregation in three-member committees with a cheap-talk communication stage followed by voting under either majority or unanimity rule. Using the joint distribution of messages and votes, we test Bayesian equilibrium predictions about both communication and voting. Committee decisions are substantially more accurate than Bayesian equilibrium predicts, and messages are close to fully truthful; subjects also correctly anticipate this truthfulness, which is difficult to reconcile with limited-depth accounts based on distrust of communication. At the same time, voting responds too strongly to private signals and to message profiles relative to Bayesian best responses, which is inconsistent with a pure lying-aversion explanation. We show that a parsimonious belief-distortion extension of quantal response equilibrium, in which agents overweight the latest information when forming posteriors, provides a unified account of communication and voting across treatments and captures most of the systematic variation in observed choice frequencies.

Keywords: committees, incomplete information, cheap talk, information aggregation, laboratory experiment, belief distortions, Bayesian updating, quantal response equilibrium

JEL Classification Codes: D71, D72, C90.

*Thanks to Guillaume Fréchette, Steffen Huck, Rune Midjord and Tomás Rodríguez Barraquer for their helpful comments and suggestions. Financial support of the WZB Berlin and the DFG (project BR 4648/1 and CRC TRR 190) is greatly appreciated. Corresponding author: Yves Breitmoser.

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

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

1 Introduction

> I've searched all the parks in all the cities and found no statues of committees. — Quote attributed to Gilbert K. Chesterton

Committees are ubiquitous decision-making institutions. They shape monetary policy (Federal Reserve Board), the safety assessment of pharmaceutical drugs (FDA advisory committees), and verdicts in criminal trials (juries). A classical argument for committees, formalized by de Condorcet (1785), is that they aggregate dispersed private information and thereby outperform any single decision maker.

This optimistic view is fragile in standard theory. Beginning with Austen-Smith and Banks (1996) and Feddersen and Pesendorfer (1996, 1997), a large literature shows that even in common-value environments, incentives need not support sincere voting or efficient information aggregation once one departs from knife-edge conditions.[1] In contrast, experimental evidence often finds that committees work surprisingly well: communication is more informative and voting is less strategic than Bayesian equilibrium would suggest, and resulting decisions are unusually accurate.[2] A central question is why committee behavior appears systematically closer to "truthful communication and sincere voting" than standard strategic models predict, and what this implies for modeling deliberation and information use in collective choice.

We study this question in a laboratory committee environment with private information, a cheap-talk communication stage, and subsequent voting under either majority or unanimity rule. Each subject receives a private signal about the state of the world, sends a binary message, observes the messages of the other committee members, and then votes. Payoffs combine a common-value component that rewards the committee for matching the state with an expressive component that rewards the individual for voting for a particular option. The expressive motive creates sharp incentives to misrepresent private information and to vote strategically, and thus provides a stringent test of the empirical prevalence of non-strategic communication and voting.

The experimental design and data are shared with a companion paper (Breitmoser and Valasek, 2024), which studies the institutional implications of majority versus unanimity in the presence of expressive incentives. The present paper instead takes the strategic environment as given and uses the *joint* distribution of messages and votes to diagnose which departures from Bayesian equilibrium are empirically useful for organizing committee behavior. The two-stage structure delivers sharp cross-stage restrictions: many candidate forces can rationalize high truth-telling in isolation, but they make distinct predictions for how subjects should use message profiles when voting, precisely because expressive incentives make strategic considerations salient at both stages. Our contribution is therefore methodological as well as substantive: we show how to use voting behavior conditional

[1] Information aggregation can fail for reasons including the decision rule (Feddersen and Pesendorfer, 1998), correlated errors (Palley and Soll, 2019), uncertainty about the signal structure (Mandler, 2012), departures from preference monotonicity (Bhattacharya, 2013), uncertainty about the size of the electorate (Ekmekci and Lauermann, 2019), or additional payoff components such as bribery (Dal Bó, 2007), preferences for winning (Callander, 2007, 2008), moral payoffs (Feddersen et al., 2009), partisan expressive motives (Morgan and Várdy, 2012), or accountability for individual votes (Midjord et al., 2017, 2021).

[2] For a seminal reference, see Goeree and Yariv (2011).

on communication to discipline behavioral hypotheses about belief formation and information use in committees.

Our main empirical findings are straightforward. First, communication is close to fully truthful even in information sets where Bayesian equilibrium predicts substantial strategic misreporting. Second, subjects correctly anticipate this truthfulness. This combination is difficult to reconcile with limited-depth accounts that rely on systematic distrust of messages. Third, voting responds too strongly to both private signals and observed message profiles relative to Bayesian best responses given the estimated degree of truth-telling. This excessive responsiveness is inconsistent not only with Bayesian equilibrium, but also with a pure "lying-aversion" explanation that fixes communication while leaving Bayesian best-response voting intact. Taken together, the data suggest that committees are accurate because members communicate honestly and then behave *as if* the informational content of signals and messages were higher than under the objective signal structure.

Motivated by extensive evidence that individuals deviate from Bayesian updating, we analyze a parsimonious belief distortion that overweights new information when forming posteriors (recent discussions include Bordalo et al., 2020; Afrouzi et al., 2023). In our setting, such overweighting implies that a subject places too much weight on her own signal when messaging and, conditional on largely truthful communication, too much weight on the public message profile when voting. The same mechanism can therefore account for both unusually truthful communication and unusually strong responsiveness of votes to message majorities. We embed this distortion in a simple structural framework by extending logit quantal response equilibrium, which allows us to use the full set of observed message and vote frequencies to quantify how much systematic variation is captured by each behavioral component.

A natural alternative interpretation is that subjects underweight base rates, effectively treating the prior as less salient than sample information. In our framework this concern is closely related to the overweighting channel: systematic downweighting of the prior can be represented as overweighting likelihood information. The data are informative because voting behavior conditional on message profiles pins down the implied mapping from public information to posterior beliefs. In the empirical and structural sections we therefore report diagnostics that translate observed voting frequencies into implied posteriors and compare these to Bayesian benchmarks. This comparison distinguishes between near-complete prior neglect and a more moderate but systematic overweighting of new information, and it clarifies which dimensions of non-Bayesian belief formation are needed to organize behavior in a strategic committee environment.

The paper makes three contributions. First, we document that committee decisions are substantially more accurate than Bayesian equilibrium predicts in a setting with strong incentives for strategic communication and voting, and we show that this accuracy reflects both high truth-telling and strong reliance on message majorities. Second, we use the two-stage structure of the game to distinguish between explanations that operate primarily through communication (such as lying aversion) and those that operate through belief formation and information use at the voting stage; the key identifying content comes from the joint restrictions linking message frequencies to subsequent votes. Third, we provide a parsimonious structural account based on a one-parameter belief distor-

tion layered onto logit QRE, and we quantify how much of the observed choice-frequency variation it captures. Beyond the laboratory, our findings speak to a broader modeling lesson: allowing for disciplined departures from Bayesian information processing can be essential for explaining why deliberative bodies sometimes appear to aggregate information effectively even when standard strategic incentives point in the opposite direction.

The remainder of the paper is organized as follows. Section 2 presents the model and the experimental design. Section 3 reports the experimental results and documents the central empirical regularities in communication and voting. Section 4 provides the structural analysis and quantifies the relative explanatory power of the candidate behavioral components. Section 5 concludes and discusses natural targets for future experimental designs aimed at sharper causal identification of belief distortions in committee settings.

1.1 Related literature

Our experiment builds on the experimental literature on information aggregation in committees and juries. Early laboratory studies analyze voting under different decision rules without pre-play communication (Guarnaschelli et al., 2000; Ali et al., 2008). Allowing for communication prior to voting, Goeree and Yariv (2011) study collective deliberation and document what they term "overcommunication": messages are more informative than equilibrium incentives would suggest, and committees often reach accurate decisions. Related evidence with heterogeneous preferences is provided by Le Quement and Marcin (2020). Our environment adds an explicit expressive voting motive, which sharpens strategic incentives and connects to experimental work that studies expressive payoffs directly (Ginzburg et al., 2022). Relative to this literature, the distinctive feature of our analysis is that we use the *joint* distribution of communication and subsequent voting to discipline a behavioral account of why committees work unexpectedly well: vote choices conditional on message profiles reveal how subjects map public deliberation into posterior beliefs and actions.

A second strand concerns why communication in cheap-talk settings is often more truthful than benchmark equilibrium predicts. Outside committees, experiments in sender–receiver and strategic information transmission games document systematic truth-telling and receiver belief-in-the-message behavior that is difficult to reconcile with purely strategic accounts (Gneezy, 2005; Charness and Dufwenberg, 2006; Cai and Wang, 2006; Sánchez-Pagés and Vorsatz, 2007). Prominent explanations emphasize non-standard preferences over messages (lying aversion and related social-image or guilt motives) or limited strategic reasoning. In committee environments, such forces can rationalize unusually truthful messages, but they are less directly informative about how subjects process message profiles when voting. Our two-stage setting turns this observation into an empirical lever: conditional voting behavior provides restrictions that separate "communication-only" explanations from models in which belief formation and information use depart from Bayesian benchmarks.

The present paper focuses on belief formation and information use after communication. A large body of evidence in psychology and economics documents systematic deviations from Bayesian updating, including base-rate neglect and excessive responsiveness to recent signals (Kahneman and

Tversky, 1973; Bar-Hillel, 1980; Tversky and Kahneman, 1982). Recent work in economics develops formal and empirical accounts of overreaction to information in applied contexts (Bordalo et al., 2020; Afrouzi et al., 2023). Our aim is not to adjudicate among the many proposed non-Bayesian updating models, but to capture, in a parsimonious way, the direction and magnitude of responsiveness to new information in a strategic committee setting. To do so, we introduce a simple belief-distortion parameter that scales the weight placed on signals and messages relative to Bayesian benchmarks, and we embed this distortion in a logit quantal response framework for extensive-form games (McKelvey and Palfrey, 1998). This allows us to use both messaging and voting frequencies, and the cross-stage restrictions linking them, to quantify how much systematic variation in behavior is explained by a single reduced-form deviation in belief updating.

Finally, we emphasize that our experiment was not tailored to discriminate sharply among competing non-Bayesian updating models. Axiomatic and structural approaches to non-Bayesian updating offer a rich set of alternatives (Epstein, 2006; Sandroni et al., 2008; Ortoleva, 2012; Massari, 2021; De Filippis et al., 2022). Distinguishing among them would naturally call for additional experimental levers, for example varying priors or the information structure to generate environments in which different forms of under- versus overreaction yield contrasting predictions. In the absence of such levers, our contribution is to show that a single-parameter belief distortion, layered onto a standard stochastic best-response model, provides a disciplined and empirically successful account of the joint patterns of truth-telling and voting in committees with expressive incentives. We view this as complementary to the companion paper (Breitmoser and Valasek, 2024): that paper asks how voting rules perform under expressive incentives, whereas we use the same environment to measure how agents depart from Bayesian information processing in a way that matters for collective choice.

2 Theory and Experimental Design

2.1 The voting game

We study a three-member Condorcet committee with a pre-vote cheap-talk stage and expressive payoffs. Nature draws a state $\omega \in \{R(ed), B(lue)\}$. A committee of $N = 3$ experts $i \in \{1,2,3\}$ receives conditionally independent private signals $s_i \in \{R,B\}$ with accuracy

$$\Pr(s_i = \omega \mid \omega) = \alpha, \qquad \alpha \in (1/2, 1), \tag{1}$$

i.i.d. across experts conditional on ω. In the experiment we set $\alpha = 0.6$. All experts share a common prior $p_0 = \Pr(\omega = R)$, which is uninformative in the experiment: $p_0 = 1/2$.

After observing s_i, each expert sends a binary message $m_i \in \{R,B\}$. After observing the message profile $m = (m_1, m_2, m_3)$, each expert simultaneously submits a vote $v_i \in \{R,B\}$ (no abstention). A voting rule $D \in \{\text{Majority}, \text{Unanimity}\}$ maps the vote profile $v = (v_1, v_2, v_3)$ into a committee decision $X \in \{R,B\}$.

Payoffs combine a common-value component and an expressive component. Let $C > 0$ denote

the payoff from a correct committee decision and let $K \in (0, C)$ denote the expressive payoff from voting for R, irrespective of the state and the committee decision. Terminal payoffs are

$$\pi_i(X, \omega, v_i) \;=\; C \cdot \mathbf{1}\{X = \omega\} \;+\; K \cdot \mathbf{1}\{v_i = R\}. \tag{2}$$

We assume risk-neutral preferences.[3]

The timing is: (i) Nature draws ω and signals $(s_i)_{i=1}^3$; (ii) experts observe s_i and simultaneously send messages $(m_i)_{i=1}^3$; (iii) experts observe m and simultaneously submit votes $(v_i)_{i=1}^3$; (iv) votes are aggregated and payoffs accrue.

Under *Majority*, the committee chooses B if and only if at least two experts vote B, and chooses R otherwise. Under *Unanimity*, our experiment allows multiple rounds to reach unanimity.[4] For theoretical exposition we follow Breitmoser and Valasek (2024) and use a reduced-form "default-R" convention: the committee chooses B if and only if all experts vote B; otherwise the committee chooses R. (In particular, any failure to attain unanimity for B is treated as selecting R.)

We restrict attention to symmetric (behavioral) strategies. A symmetric strategy profile is a pair (σ, τ) where $\sigma : \{R, B\} \to [0, 1]$ is a messaging rule with $\sigma(s)$ equal to the probability of sending message R after signal s, and τ is a voting rule with $\tau(s, m_i, M)$ equal to the probability of voting R after private signal s, own message m_i, and message-profile summary $M = \#\{j : m_j = B\}$ (the total number of B-messages in the committee).[5] The expressive payoff K generates a collective-action problem: when an expert is unlikely to be pivotal, she strictly prefers voting R regardless of information. This tension produces strategic incentives both in messaging and in voting and is central for distinguishing behavioral explanations.

2.2 Behavioral models and predictions

This section fixes a small set of behavioral models and records the corresponding *point predictions* used in the empirical analysis. The models are designed to perturb one component of standard strategic behavior at a time: belief formation (*overreaction*), preferences over messages (*lying aversion*), or strategic reasoning (*level-k*). In every case we impose symmetry and sequential optimality *given the model's beliefs and preferences*. When a model admits multiple symmetric predictions, we select a single point prediction using the limiting-logit procedure formalized in Appendix A.4 (Definition 10) and implemented numerically in Appendix A.7. The role of this selection is purely to make the quantitative comparisons in Tables 1–2 unambiguous.

Baseline concepts. We begin by fixing the benchmark equilibrium notion and the refinement used when we need a unique point prediction.

[3] With risk aversion, the risky common-value term effectively receives less weight relative to the sure expressive term, strengthening incentives to vote R and thereby working against information aggregation.

[4] This approximates deliberation via straw polls (see Guarnaschelli et al., 2000; Goeree and Yariv, 2011).

[5] Given m_i and M, expert i can infer the number of B-messages sent by the other two experts.

Table 1: Summary of theoretical predictions: Majority

	Messages		Voting			
	$s_i = R$	$s_i = B$	$M = 0$	$M = 1$	$M = 2$	$M = 3$
Majority 40-10						
Equilibrium	✓	×	✓	✓	✓	×
Lying Aversion	✓	✓	✓	✓	×	✓
Overreaction	✓	×	✓	✓	✓	✓
Level-K (1)	×	×	✓	✓	×	×
Level-L (1)	✓	✓	✓	✓	×	✓
Altruism	×	×	✓	✓	✓	✓
Majority 35-15						
Equilibrium	×	×	✓	✓	×	×
Lying Aversion	✓	✓	✓	✓	×	×
Overreaction	✓	×	✓	✓	✓	✓
Level-K (1)	×	×	✓	✓	×	×
Level-L (1)	✓	✓	✓	✓	×	×
Altruism	×	×	✓	✓	✓	✓

Note: For the overreaction row in Tables 1–2, we report the limit $\kappa \to \infty$ ("perfect overreaction") to highlight sharp qualitative contrasts. The Appendix reports the general κ case and the numerical method used to compute predictions.
For *Messages*: ✓ indicates truthful message, and × indicates non-truthful message. For *Voting*: ✓ indicates that the committee votes with the majority of messages with over 50% probability, and × indicates that the committee votes with the majority of messages with under 50% probability. Highlighted cells indicate that predictions differ from "Equilibrium."

Table 2: Summary of theoretical predictions: Unanimity

	Messages		Voting			
	$s_i = R$	$s_i = B$	$S = 0$	$S = 1$	$S = 2$	$S = 3$
Unanimity 40-10						
Equilibrium	×	×	✓	✓	×	✓
Lying Aversion	✓	✓	✓	✓	×	✓
Overreaction	✓	✓	✓	✓	✓	✓
Level-K (1)	×	×	✓	✓	×	×
Level-L (1)	✓	✓	✓	✓	×	✓
Altruism	✓	✓	✓	✓	✓	✓
Unanimity 35-15						
Equilibrium	×	×	✓	✓	×	✓
Lying Aversion	✓	✓	✓	✓	×	✓
Overreaction	✓	✓	✓	✓	✓	✓
Level-K (1)	×	×	✓	✓	×	×
Level-L (1)	✓	✓	✓	✓	×	✓
Altruism	✓	✓	✓	✓	✓	✓

Note: For *Messages*: ✓ indicates truthful message, and × indicates non-truthful message. For *Voting*: ✓ indicates that the committee votes with the majority of signals with over 50% probability, and × indicates that the committee votes with the majority of signals with under 50% probability. Highlighted cells indicate that predictions differ from "Equilibrium."

Definition 1 (Benchmark equilibrium)**.** *Our benchmark notion is* symmetric sequential equilibrium *of* Γ *with Bayesian beliefs.*

For some parameterizations, symmetric sequential equilibrium is not single-valued, and certain behavioral benchmarks (e.g. distorted beliefs combined with stochastic choice) generate a family of

admissible predictions indexed by a "precision" parameter. We therefore use logit quantal response as a disciplined way to select a point prediction.

Definition 2 (Logit quantal response equilibrium)**.** *Fix a behavioral specification (Bayesian beliefs or distorted beliefs). A* symmetric logit QRE *is a symmetric assessment in which, at every information set I, each feasible action a is chosen with probability*

$$\Pr(a \mid I) \;=\; \frac{\exp\{\lambda EU(a \mid I)\}}{\sum_{a' \in A(I)} \exp\{\lambda EU(a' \mid I)\}},$$

where $EU(\cdot \mid I)$ denotes the action's continuation payoff under the assessment and $\lambda > 0$ is a payoff-sensitivity parameter. When we require a single point prediction, we use the limiting-logit selection as $\lambda \to \infty$ (Appendix A.4, Definition 10).

Preference- and payoff-based benchmarks. The next two benchmarks shut down a specific strategic force by assumption: lying aversion eliminates message manipulation, and altruism eliminates the expressive motive.

Definition 3 (Lying aversion)**.** *The lying-aversion benchmark fixes truthful messaging: $\sigma(R) = 1$ and $\sigma(B) = 0$. Voting is then chosen to be sequentially rational given Bayesian beliefs and the induced informativeness of messages.*

Definition 4 (Altruism)**.** *The altruism benchmark sets the expressive motive aside: agents behave as if $K = 0$, equivalently choosing strategies that maximize $\Pr(X = \omega)$.*

Overreaction as a belief distortion. Our main belief-based deviation is a one-parameter distortion of Bayesian posteriors. The distortion is chosen to (i) move posteriors monotonically in the direction of the Bayesian posterior and (ii) scale log-odds by a single parameter, which yields transparent comparative statics. Let I denote any information set (after observing the private signal, or after observing the message profile). Write

$$q(I) \;=\; \Pr(\omega = B \mid I)$$

for the Bayesian posterior under the true signal structure and the strategy profile under consideration. For $\kappa > 0$, define

$$T_\kappa(q) \;=\; \frac{q^\kappa}{q^\kappa + (1-q)^\kappa}, \qquad q \in [0,1], \tag{3}$$

so that posterior odds satisfy $\frac{T_\kappa(q)}{1-T_\kappa(q)} = \left(\frac{q}{1-q}\right)^\kappa$. The Bayesian benchmark corresponds to $\kappa = 1$. Values $\kappa > 1$ generate *overreaction* (posteriors become more extreme), while $\kappa \in (0,1)$ would correspond to underreaction.

Definition 5 (κ-overreaction)**.** *Fix $\kappa \geq 1$. In the κ-overreaction model, at every information set I, agents evaluate expected payoffs using the distorted posterior $\hat{q}_\kappa(I) = T_\kappa(q(I))$ in place of the Bayesian posterior $q(I)$. The mapping T_κ and parameter κ are common knowledge.*

The "perfect overreaction" benchmark corresponds to $\kappa \to \infty$; in that limit any Bayesian posterior strictly above (below) $1/2$ is perceived as certainty for B (R).

The next lemma collects two properties used repeatedly below: T_κ is monotone in the Bayesian posterior, and for posteriors that already favor B (resp. R), increasing κ pushes beliefs further toward certainty for B (resp. R).

Lemma 1 (Extremeness and the limit case)**.** *For any $q \in (0,1)$, $T_\kappa(q)$ is strictly increasing in q. Moreover, if $q > 1/2$, then $T_\kappa(q)$ is strictly increasing in κ and $\lim_{\kappa\to\infty} T_\kappa(q) = 1$; if $q < 1/2$, it is strictly decreasing in κ and $\lim_{\kappa\to\infty} T_\kappa(q) = 0$.*

Proof. Monotonicity in q follows from differentiating (3) or from the odds representation. For the comparative statics in κ, note that

$$\frac{T_\kappa(q)}{1-T_\kappa(q)} = \left(\frac{q}{1-q}\right)^\kappa.$$

If $q > 1/2$ then $\frac{q}{1-q} > 1$, so the right-hand side is strictly increasing in κ and diverges to $+\infty$ as $\kappa \to \infty$, implying $T_\kappa(q) \uparrow 1$. If $q < 1/2$ then $\frac{q}{1-q} < 1$, so the right-hand side is strictly decreasing in κ and converges to 0, implying $T_\kappa(q) \downarrow 0$. □

Interpretation. With prior $1/2$, the Bayesian posterior after a private signal equals α for the realized signal. For $\kappa > 1$, the agent behaves as if signals (and likewise informative message profiles) were more precise, since $T_\kappa(\alpha) > \alpha$ and, more generally, $T_\kappa(q)$ lies farther from $1/2$ than q whenever $q \neq 1/2$.

Level-k benchmarks. We also consider two bounded-reasoning benchmarks that modify strategic reasoning rather than preferences or beliefs. These are not equilibrium concepts: each agent is assumed to be sequentially optimal given a misspecified belief about opponents' play.

Definition 6 (Level-k)**.** *In the level-k benchmark with $k = 1$, agents best respond at both stages to the belief that opponents randomize uniformly, i.e. they believe $\sigma(\cdot) = 1/2$ and $\tau(\cdot,\cdot,\cdot) = 1/2$.*

Definition 7 (Level-ℓ)**.** *In the level-ℓ benchmark with $\ell = 1$, agents best respond to the belief that opponents communicate truthfully and randomize uniformly when voting, i.e. they believe $\sigma(R) = 1$, $\sigma(B) = 0$, and $\tau(\cdot,\cdot,\cdot) = 1/2$.*

A key voting-stage implication

A central identifying contrast in the paper is that, holding (near) truthful communication fixed, overreaction predicts *amplified* responsiveness to message profiles at the voting stage relative to Bayesian best responses. The reason is simple: under Majority, voting R yields the expressive payoff regardless of whether the vote is pivotal, whereas voting B sacrifices that payoff and is rewarded only through

the event of being pivotal. Appendix A.4 formalizes this "pivotality times posterior advantage" decomposition (Lemma 2). The proposition below states the corresponding voting threshold under κ-overreaction and highlights the comparative statics in κ.

Fix Majority and a voting-stage information set I (after observing s_i, the own message, and the message profile). Let $q(I) = \Pr(\omega = B \mid I)$ be the Bayesian posterior, $\hat{q}_\kappa(I) = T_\kappa(q(I))$ the distorted posterior, and let $\pi(I) \in (0,1]$ denote the probability that expert i is pivotal at I given opponents' strategies.

Proposition 1 (Overreaction amplifies responsiveness in voting)**.** *Fix an information set I with $\pi(I) > 0$. Under κ-overreaction, voting B is optimal at I if and only if*

$$\hat{q}_\kappa(I) \;\geq\; \frac{1}{2} + \frac{K}{2\pi(I)C}. \tag{4}$$

Moreover, whenever $q(I) > 1/2$, the left-hand side $\hat{q}_\kappa(I) = T_\kappa(q(I))$ is increasing in κ, so (4) is (weakly) easier to satisfy for $\kappa > 1$ than for $\kappa = 1$. Thus, conditional on information that favors B, overreaction weakly increases the propensity to vote for B.

Proof. Voting R yields the expressive payoff K regardless of whether the vote is pivotal. The common-value payoff depends on the vote only if the agent is pivotal: conditional on pivotality, voting B yields expected common-value payoff $C\hat{q}_\kappa(I)$, while voting R yields $C(1-\hat{q}_\kappa(I))$. Hence the expected gain from voting B rather than R equals

$$\pi(I)C\big(2\hat{q}_\kappa(I) - 1\big) - K.$$

Voting B is optimal if and only if this gain is nonnegative, which is equivalent to (4). If $q(I) > 1/2$, then Lemma 1 implies $\hat{q}_\kappa(I)$ is increasing in κ, so the inequality becomes weakly easier to satisfy as κ rises. □

Overview predictions and experimental parameterization

The experiment varies the strength of expressive incentives. The *low* expressive-payoff treatment uses $(\alpha, K, C) = (0.6, 10, 40)$ and the *high* expressive-payoff treatment uses $(\alpha, K, C) = (0.6, 15, 35)$. These parameterizations change the welfare trade-off between inducing votes that support information aggregation and the private benefit from voting R. In particular, in the low treatment the efficient committee action conditional on the signal profile coincides with following the (posterior) majority of signals, whereas in the high treatment it is efficient to select B only when all three signals are B.

Tables 1–2 report predicted choice frequencies under each behavioral model for these experimental parameters. For overreaction we also report the limit $\kappa \to \infty$ ("perfect overreaction") to highlight sharp qualitative contrasts. Appendix A.7 reports the numerical procedure and the full predicted strategy components for general κ. Several models admit multiple symmetric equilibria under these parameters, especially under Unanimity (where B requires unanimous B votes). In those cases we apply the limiting-logit selection in Definition 2 (see Appendix A.4, Definition 10).

At a high level, the comparative statics under Majority are as follows. In the benchmark equilibrium, communication is strategically distorted and voting is weakly responsive to informative message profiles because the expressive payoff depresses the incentive to be pivotal for B (see Appendix A.4, Lemma 2 for the underlying voting decomposition, and Propositions 3–2 for the parameterized benchmark predictions). Lying aversion fixes communication to be truthful but leaves the pivotality logic intact, so it predicts limited voting with the informational majority even when messages reveal strong evidence for B (Appendix A.4, Proposition 4). Overreaction instead distorts posteriors in a direction that makes informative evidence feel more decisive, thereby increasing both the propensity to report in line with one's signal and the propensity to vote with the informational majority of messages (Proposition 1, together with the monotone comparative statics in Appendix A.4, Lemma 6). Under the bounded-reasoning benchmarks, level-k predicts babbling communication together with voting dominated by the expressive motive, while level-ℓ inherits truthful communication by assumption but implies weak use of message profiles at the voting stage.

We next provide additional detail for our three main concepts—Benchmark equilibrium, Lying aversion, and Overreaction—for the Majority rule. Throughout, recall that $\tau(s,m,M)$ denotes the probability of voting R (so $\tau = 0$ corresponds to voting B with certainty).

Benchmark equilibrium (Majority). We begin with symmetric sequential equilibrium under Bayesian beliefs. Under Majority, the expressive payoff is earned whenever one votes R, while the vote affects the committee outcome only through pivotality. The resulting collective-action problem is most transparent at information sets where public information strongly favors B: if experts were to communicate truthfully and then vote mechanically with the message majority, then in those profiles each expert would anticipate that the others vote B with high probability, making her rarely pivotal; the sure expressive payoff would then make deviation to R attractive. Anticipating weak voting incentives precisely in the most informative profiles, equilibrium messaging becomes strategic: B-types sometimes send R to reduce the frequency with which they (and others) would face the trade-off between forgoing K and inducing an accurate B decision.

Under the experimental high expressive-payoff calibration, this force can fully unravel aggregation: there exists a symmetric sequential equilibrium in which all players vote R at every voting information set, rendering messages payoff-irrelevant on path (Appendix A.4, Proposition 2). Under the low expressive-payoff calibration, we report the limiting-logit Bayesian point prediction (Appendix A.4, Proposition 3).

Prediction 1 (Equilibrium: Majority)**.** *For low expressive payoffs, the limiting-logit Bayesian point prediction has* $\sigma(R) = 1$*,* $\sigma(B) = 0.56$ *and experts vote for B if and only if* $M = 2$ *and* $s_i = B$ *(i.e.* $\tau(B,B,2) = 0$*, and* $\tau(\cdot,\cdot,\cdot) = 1$ *otherwise). For high expressive payoffs, experts babble in the message stage (*$\sigma(R) = \sigma(B) = 0.5$*) and vote R for all message profiles (*$\tau(\cdot,\cdot,\cdot) = 1$*). These values correspond to Appendix A.4, Propositions 3–2, with computation in Appendix A.7.*

Lying aversion (Majority). Lying aversion fixes truthful messaging by assumption, $\sigma(R) = 1$ and $\sigma(B) = 0$, thereby eliminating strategic manipulation at the communication stage. The voting stage

nevertheless inherits the pivotality logic. With truthful messages, the message profile reveals the signal profile, so the posterior in favor of B is maximized at (B,B,B). Even there, voting B need not be optimal with probability one: if one expects the other two experts to vote B with sufficiently high probability, pivotality is sufficiently unlikely that the sure expressive payoff from voting R can dominate. Appendix A.4, Proposition 4 provides the exact implication under our experimental calibrations, including the unique symmetric mixed equilibrium at *BBB* in the low treatment.

Prediction 2 (Lying Aversion: Majority)**.** *By definition, experts communicate truthfully ($\sigma(R) = 1$, $\sigma(B) = 0$). For low expressive payoffs, experts vote for B only if $M = 3$, and at $M = 3$ they vote R with probability $\tau(B,B,3) = 1 - q^\star \approx 0.3595$ (equivalently, vote B with probability $q^\star \approx 0.6405$), while $\tau(\cdot,\cdot,\cdot) = 1$ otherwise. For high expressive payoffs, experts vote R for all message profiles ($\tau(\cdot,\cdot,\cdot) = 1$). See Appendix A.4, Proposition 4.*

Overreaction (Majority). Overreaction modifies beliefs rather than incentives: experts remain sequentially optimal, but evaluate expected payoffs under distorted posteriors. Two implications follow.

First, at the voting stage, Proposition 1 shows that the decision rule depends on $\hat{q}_\kappa(I) = T_\kappa(q(I))$. Compared to $\kappa = 1$, a larger κ increases $\hat{q}_\kappa(I)$ whenever the Bayesian posterior favors B, making it easier to justify sacrificing the expressive payoff in favor of an accurate B decision. In this sense overreaction counteracts the pivotality problem by making the perceived stakes of choosing the wrong alternative larger.

Second, at the messaging stage, overreaction makes private information feel more decisive. With prior $1/2$, an expert who observes s_i has Bayesian posterior $\Pr(\omega = s_i \mid s_i) = \alpha$, but under overreaction she perceives $T_\kappa(\alpha) > \alpha$ (Lemma 1). The perceived expected loss from inducing an incorrect committee decision through a strategic misreport therefore increases. Moreover, because overreaction raises voting-stage responsiveness, messages are perceived as more consequential for final outcomes, which further strengthens incentives to report in line with the signal.

Prediction 3 (Overreaction: Majority)**.** *With overreaction, experts communicate strategically after signal B, but at a much lower rate than under Equilibrium ($\sigma(R) = 1$; low expressive payoffs $\sigma(B) = 0.10$; high expressive payoffs $\sigma(B) = 0.28$). For both low and high expressive payoffs, experts who message B vote for B with certainty if $M = 2$ and with positive probability if $M = 3$ (i.e. $\tau(B,B,2) = 0$; low expressive payoffs $\tau(B,B,3) = 0.15$; high expressive payoffs $\tau(B,B,3) = 0.31$). These values correspond to the limiting-logit point prediction for the distorted-belief logit-QRE (Definitions 2 and 5); see Appendix A.7 for the computation.*

Comparing Lying aversion and Overreaction clarifies why the voting stage is essential for discrimination. Both concepts can accommodate highly truthful messages. However, with Bayesian voting (lying aversion), pivotality keeps the incentive to vote R strong exactly in profiles where others are expected to vote B, so voting with the informational majority remains limited (Appendix A.4, Proposition 4). With overreaction, distorted beliefs raise $\hat{q}_\kappa(I)$ precisely in those profiles, and Proposition 1 implies a higher propensity to vote with the informational majority of messages, especially

when expressive incentives would otherwise induce strategic R voting.

2.3 Experimental design

The experimental design is shared with a companion paper that studies institutional implications of majority versus unanimity voting in the presence of expressive incentives (Breitmoser and Valasek, 2024). The present paper takes the environment as given and uses the resulting joint patterns of messages and votes to discipline behavioral explanations. The identifying variation is a 2×2 between-subjects design that varies (i) the decision rule (Majority vs. Unanimity) and (ii) the strength of expressive incentives (Low vs. High). The decision rule shifts pivotality and therefore strategic voting incentives; the expressive incentive shifts the severity of the collective-action problem. Together, these variations generate sharp differences in predicted behavior across communication and voting, which we exploit throughout the theoretical and empirical analysis.

The experiment closely follows Guarnaschelli et al. (2000) and Goeree and Yariv (2011). We use neutral language throughout, represent uncertainty via urns and ball draws, and provide feedback on the realized state and payoffs after each game. The sessions were conducted at the WZB/TU experimental laboratory in Berlin in May, June, and November 2016. Subjects were recruited using ORSEE (Greiner, 2015) and the experiment was programmed in z-Tree (Fischbacher, 2007). A translation of the instructions and a composite screenshot are provided in Appendix B.

Treatments and payoffs. Table 3 summarizes the four treatments. In all treatments, the common-value payoff C (earned if the committee decision matches the true state) and the expressive payoff K (earned if the subject votes for R) satisfy $C+K=50$ points. In the Low treatments, $(C,K)=(40,10)$; in the High treatments, $(C,K)=(35,15)$. The calibration is chosen so that the incentive to vote expressively remains salient even in information sets with relatively favorable public information for B (as discussed in Section 2.2). Signal precision is constant across treatments and equal to $\alpha=0.6$.

Table 3: Overview of experimental treatments

Label	Decision rule	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

Subjects were paid the sum of points accumulated across all 50 games; one point corresponded to one euro cent in all treatments. Sessions lasted between 75 and 105 minutes, and average earnings were between 19 and 22 Euros across sessions.

Information and salience. Each game begins with an explicit presentation of uncertainty and incentives. The prior over states is fixed and uninformative ($p_0 = 1/2$), and the signal structure is conveyed using urns and ball draws, with the instructions stating the signal accuracy ($\alpha = 0.6$). The

payoff parameters (C, K) are displayed in points and remain constant within a session. These features are made salient at the time of decision making: subjects observe their private signal before sending a message, and the payoff consequences of (i) matching the true state and (ii) voting for R are shown in the interface and reiterated in the instructions and control questions. This matters for interpretation later, because it limits the scope for explanations based purely on misunderstanding of the prior or the signal structure.

Procedures. We ran 16 sessions in total (four per treatment). Fourteen sessions had 12 participants and two sessions had 9 participants; participants were seated randomly upon arrival. An experimental assistant distributed printed instructions and read them aloud. Subjects then completed computerized control questions; the experiment did not start until all participants answered all questions correctly.

Subjects played 50 games in committees of size three ($N = 3$), with random rematching after each game. After each game, subjects received feedback on the realized state, the committee decision, and their payoff. The feedback screen also reported the realized signal profile and aggregate behavior in the session.

Under Majority, each game proceeded as follows: subjects observed their private signal $s_i \in \{R, B\}$, simultaneously sent a public message $m_i \in \{R, B\}$, observed the message profile, and then simultaneously submitted a vote $v_i \in \{R, B\}$. Under Unanimity, the timing was identical up to the voting stage, but subjects were given up to three voting attempts to reach unanimity. If unanimity was not reached after the third vote, the game ended with a default committee decision of R (equivalently, all subjects were assigned a default vote of R).

Upon completion of the experiment, subjects were paid individually in a separate room.

Identifying moments. The design generates a small set of particularly informative conditional frequencies because it produces sharp changes in (i) incentives to misreport in the messaging stage and (ii) incentives to vote against information in the voting stage, while holding fixed the signal structure and the salience of payoffs. First, the expressive payoff K creates information sets in which strategic considerations predict systematic distortions in communication: when an expert's signal is B, sending message R can be privately valuable because it reduces the likelihood that the committee selects B in profiles where an individual would otherwise be tempted to vote R for expressive reasons. Hence, the frequency of $m_i = R$ conditional on $s_i = B$ (and how this frequency changes between the Low and High treatments) is an informative measure of the extent to which subjects respond strategically to the collective-action problem at the communication stage.

Second, conditional on a given message profile, the voting rule changes pivotality and therefore the strength and location of incentives to vote expressively. Under Majority, pivotality is concentrated in knife-edge profiles; under Unanimity with a default-R outcome, the committee selects B only if all votes are B, which changes both the pivotal events and the mapping from vote profiles to outcomes. As a result, the same observed message profile can imply different strategic incentives across rules even when beliefs are held fixed. Comparing vote frequencies across Majority and Unanimity, conditioning on message profiles and the subject's own signal, therefore provides additional restrictions

beyond what message truthfulness alone can deliver.

Third, the two stages jointly tighten identification. Many explanations can rationalize high truth-telling in isolation (for example, a preference for honesty), but they differ in their implications for voting once truth-telling is anticipated. In particular, if messages are believed to be highly informative, then Bayesian best-response voting with expressive incentives predicts substantial strategic R voting in message profiles that favor B, because pivotality is limited. By contrast, belief distortions that overweight new information predict systematically stronger responsiveness to message profiles in the voting stage. Accordingly, the empirical object that disciplines the behavioral analysis is the joint distribution of (m_i, v_i) conditional on (s_i, M) across treatments, rather than any single behavioral margin.

3 Empirical Analysis

We organize the empirical analysis around four questions. First, are committee decisions as accurate as predicted by the benchmark models? Second, is communication consistent with the strategic and behavioral incentives embedded in those models? Third, conditional on communication, is voting consistent with (Bayesian or distorted) best responses? Fourth, which deviations from the benchmark are most useful for jointly organizing messages and votes?

3.1 Are committees as efficient as predicted?

Table 4 reports average behavior by treatment. Table 5 then translates behavior into an information-aggregation metric: the relative frequency with which the committee decision coincides with the majority of private signals.[6]

The benchmark logic is straightforward. Under Majority, accurate aggregation is achieved if experts vote in line with their information when pivotal. Under Unanimity with a default-R outcome, accuracy additionally requires that informative messages translate into B votes when warranted; otherwise a single R vote (or a failure to reach unanimity) induces R. Expressive payoffs create an incentive to vote R even when information favors B, and (paradoxically) more informative communication can facilitate such expressive voting: the more precisely experts infer others' signals, the easier it is to identify information sets in which voting R secures the expressive bonus while being unlikely to change the outcome. In this sense, even a force that increases truth-telling (such as lying aversion) need not increase decision accuracy once voting incentives are taken into account.

Against this background, the most accuracy-friendly benchmark in our theoretical menu is the equilibrium selected by our ex ante payoff criterion, which predicts accuracy rates of 0.617, 0.500, 0.640, and 0.640 across treatments (Table 5). Observed accuracy is substantially higher. Across treatments, and separately for inexperienced and experienced play (first versus second half of each

[6]Because payoffs include an expressive component, this "efficiency" measure is best interpreted as an index of information aggregation rather than welfare. It is, however, the natural accuracy benchmark for a Condorcet-style environment with conditionally independent signals.

Table 4: Observed choices in the experiment

	Messages		Voting											
	$\sigma(R)$	$\sigma(B)$	$\tau(R,R,0)$	$\tau(B,R,0)$	$\tau(R,R,1)$	$\tau(R,B,1)$	$\tau(B,B,1)$	$\tau(B,R,1)$	$\tau(R,R,2)$	$\tau(R,B,2)$	$\tau(B,B,2)$	$\tau(B,R,2)$	$\tau(R,B,3)$	$\tau(B,B,3)$
Observations across treatments														
Majority 40-10	0.9	0.16	0.98	0.77	0.94	0.86	0.92	0.82	0.64	0.68	0.44	0.51	0.55	0.31
	(0.01)	(0.01)	(0.01)	(0.06)	(0.01)	(0.06)	(0.02)	(0.04)	(0.03)	(0.06)	(0.02)	(0.08)	(0.11)	(0.03)
Majority 35-15	0.79	0.13	0.98	0.92	0.96	0.91	0.92	0.78	0.75	0.94	0.6	0.69	0.84	0.32
	(0.01)	(0.01)	(0.01)	(0.06)	(0.01)	(0.04)	(0.02)	(0.05)	(0.03)	(0.02)	(0.02)	(0.06)	(0.04)	(0.03)
Unanimity 40-10	0.96	0.14	0.99	0.88	0.94	0.76	0.94	0.91	0.28	0.53	0.26	0.26	0.17	0.02
	(0.01)	(0.01)	(0)	(0.05)	(0.01)	(0.11)	(0.01)	(0.03)	(0.03)	(0.12)	(0.02)	(0.08)	(0.11)	(0.01)
Unanimity 35-15	0.94	0.14	1	0.95	0.96	0.59	0.92	0.87	0.48	0.67	0.43	0.28	0.38	0.03
	(0.01)	(0.01)	(0)	(0.03)	(0.01)	(0.12)	(0.02)	(0.04)	(0.03)	(0.07)	(0.02)	(0.07)	(0.14)	(0.01)

Table 5: Efficiency across treatments in relation to predictions

	Majority		Unanimity	
	40-10	35-15	40-10	35-15
Predictions				
Exp. Payoff	0.617	0.5	0.64	0.64
Lying Aversion	0.599	0.5	0.64	0.64
Overreaction	0.92	0.789	1	1
Level-K (1)	0.5	0.5	0.5	0.5
Level-L (1)	0.64	0.5	0.64	0.5
Observations				
Aggregate	0.681	0.669	0.777	0.732
Inexperienced	0.690	0.667	0.768	0.748
Experienced	0.672	0.671	0.787	0.718

Note: The table lists predicted and observed relative frequencies of committees choosing "optimally" contingent on their aggregate private information, i.e. choosing the option corresponding to the majority of private signals.

session), the corresponding rates are 0.681, 0.669, 0.777, and 0.732. Treating each session-half as an independent observation, realized accuracy falls below the benchmark prediction in only one of the 32 session-halves (the second half of session 4 in the 35–15 Unanimity treatment). A one-sided Wilcoxon test rejects the null hypothesis that accuracy does not exceed the benchmark prediction at $p < 10^{-5}$. Committees therefore aggregate information more effectively than predicted by Bayesian equilibrium (with or without the standard behavioral modifications emphasized in this literature).

Result 1. *Committees are much more efficient in aggregating information than predicted by Bayesian equilibrium.*

Among the candidate mechanisms discussed above, overreaction and altruism are the two forces that naturally point toward *higher* accuracy than Bayesian equilibrium. Moreover, only overreaction naturally predicts a qualitative asymmetry between Majority and Unanimity in how accuracy responds to public information. Table 5 reports the predictions for the extreme case of "full overreaction," in which agents behave as if their private signals were almost surely correct (and this is common knowledge). This benchmark overshoots the data, but it points to a useful direction: the deviations from Bayesian equilibrium are consistent with an *intermediate* degree of overreaction. We next examine messages and votes in isolation, before turning to a joint (structural) account.

3.2 Do subjects communicate rationally?

We begin by analyzing communication in relation to the *empirically relevant* continuation-payoff incentives associated with messages. Table 6 reports, for each treatment and each signal realization, (i) the observed relative frequency of sending message R and (ii) the average continuation payoff observed in the experiment conditional on the sender's signal and message. That is, the "Exp Payoff" entries are sample averages over subsequent votes and outcomes *as realized in the data*, conditioning only on the sender's signal and message.

Two descriptive patterns are robust. First, messages are tightly linked to private signals. Following an R signal, subjects send R with high probability (between 0.787 and 0.958 across treatments). Following a B signal, they send R with low probability (between 0.128 and 0.163). Communication is therefore close to truthful, with slightly higher truthfulness in the Unanimity treatments.

Second, message behavior is only weakly aligned with continuation-payoff incentives. The empirical payoff advantage of messaging R rather than B varies substantially across treatments and even changes sign, yet the frequency of R messages after a B signal remains tightly clustered around 0.13–0.16. Likewise, after an R signal the payoff difference between R and B messages is small in the Majority treatments, but the probability of messaging R remains high and varies non-monotonically across those treatments. Overall, continuation-payoff differences appear to be a second-order determinant of messages relative to the private signal.

Econometric check. To assess the relationship between messages and continuation-payoff incentives more systematically, we estimate a logit model in which a subject's message responds to (estimated) continuation payoffs and to her private signal. Formally, let $EP(m = R \mid s)$ and $EP(m = B \mid s)$

Table 6: Expected payoffs and relative frequencies of messages contingent on signal

(a) Majority decisions

		Exp Payoff		
		Mess *R*	Mess *B*	Rel Freq *R*
40-10	Signal *R*	32.86	32.3	0.904
	Signal *B*	25.48	27.94	0.163
35-15	Signal *R*	33.59	33.34	0.787
	Signal *B*	27.69	26.44	0.128

(b) Unanimity decisions

		Exp Payoff		
		Mess *R*	Mess *B*	Rel Freq *R*
40-10	Signal *R*	33.66	28.12	0.958
	Signal *B*	27.19	27.37	0.14
35-15	Signal *R*	35.76	30.83	0.94
	Signal *B*	29.64	27.67	0.137

Note: Continuation payoffs observed in the experiment, conditional on the private signal being *R* or *B* and the subject sending message *R* or *B* (columns "Mess *R*" and "Mess *B*"). "Rel Freq *R*" lists the observed relative frequency of sending message *R* conditional on the signal.

denote the expected continuation payoff associated with sending message R or B after signal s. Allowing for logistic errors, and letting subjects place weight λ on continuation payoffs and weight γ on their signal, we specify

$$\Pr(m = R \mid s) = \frac{\exp\{\lambda \cdot EP(m = R \mid s) + \gamma \cdot I_{s=R}\}}{\exp\{\lambda \cdot EP(m = R \mid s) + \gamma \cdot I_{s=R}\} + \exp\{\lambda \cdot EP(m = B \mid s) + \gamma \cdot I_{s=B}\}}. \tag{5}$$

Let $dEP(s) = EP(m = B \mid s) - EP(m = R \mid s)$ denote the payoff difference between sending B and R after signal $s \in \{R, B\}$. Rearranging (5) yields the regression form

$$\Pr(m = R \mid s) \;=\; \frac{1}{1 + \exp\{\lambda \cdot dEP(s) - \tilde{\gamma} \cdot (I_{s=R} - 1/2)\}}, \tag{6}$$

where $\tilde{\gamma} = 2\gamma$ absorbs the normalization induced by $I_{s=B} = 1 - I_{s=R}$.

Two implementation details matter. First, $dEP(s)$ is estimated from the data and therefore measured with error. We estimate standard errors for $dEP(s)$ (for each s and treatment) and use the MCMC correction proposed by Hadfield (2010). Second, we include random effects at the subject level to account for the panel structure. We report estimates separately for Majority, for Unanimity, and pooled across treatments.

Table 7 confirms the descriptive impression. The own-signal coefficient is large and highly significant. The continuation-payoff term is statistically insignificant in the Majority treatments and statistically significant under Unanimity and in the pooled sample, but its magnitude is modest: even when the payoff differences in Table 6 are largest (roughly 5–6 points), the implied shift in log-odds from the payoff channel is small relative to the signal effect. We therefore conclude that messages are close to truthful and only weakly strategic in the sense of being only weakly responsive to continuation-payoff incentives.

Why overreaction can support truth-telling. In our theoretical menu (Table 14), message-stage behavior can be rationalized by either lying aversion or belief distortions (overreaction), among other possibilities. Overreaction is useful here because it makes truth-telling *instrumentally* attractive:

Table 7: Messaging regressions: continuation payoffs and private signals

	Majority	Unanimity	Pooled
Estimates			
Exp. payoff	-0.296 (0.555)	0.301^{***} (0.049)	0.291^{*} (0.121)
Own signal	5.561^{***} (0.549)	5.394^{***} (0.203)	4.871^{***} (0.224)
Additional information			
Number observations	4650	4650	9300
DIC	2835.1	1907.5	5125.1
Measurement error correction	✓	✓	✓
Random effects	✓	✓	✓

Note: Logit regression of the dependent variable $\mathbf{1}\{m = R\}$ with subject-level random effects, pooling all rounds. "Exp. payoff" refers to $dEP(s)$ in (6). DIC is the deviance information criterion. One asterisk indicates $p < 0.05$, two asterisks indicate $p < 0.01$, and three asterisks indicate $p < 0.001$.

when a subject overweights the event that $\omega = s$, she overestimates the expected payoff gain from sending the message that matches her signal.

To state this connection cleanly, let $\bar{\pi}(m \mid \omega)$ denote the average continuation payoff associated with sending message $m \in \{R, B\}$ when the true state is $\omega \in \{R, B\}$. Define the state-contingent payoff difference between sending B and sending R by

$$d\pi(\omega) \;=\; \bar{\pi}(B \mid \omega) - \bar{\pi}(R \mid \omega). \tag{7}$$

Given a posterior $\Pr(\omega = \cdot \mid s)$ after observing signal s, the subjective expected payoff difference from sending B rather than R is

$$dEP(s) \;=\; \Pr(\omega = R \mid s) \cdot d\pi(R) \;+\; \Pr(\omega = B \mid s) \cdot d\pi(B). \tag{8}$$

If $\Pr(\omega = s \mid s)$ is biased upward (overreaction), then $dEP(s)$ is biased toward the message that matches s whenever correct messages are more profitable than incorrect ones.

Table 8 shows exactly this payoff structure: conditional on the true state, sending the correct message is much more profitable than sending the wrong message in all treatments. Hence, a distorted belief that places excessive posterior weight on $\omega = s$ creates a strong perceived incentive to message truthfully.

Estimating the degree of overreaction from messages alone. To quantify how strong a belief distortion would need to be to rationalize message behavior, we parameterize beliefs about the true state after observing the private signal by

$$\Pr(\omega = R \mid s) \;=\; \frac{1}{1 + \exp\{\alpha_m \cdot (I_{s=B} - 1/2)\}} \;=\; 1 - \Pr(\omega = B \mid s), \tag{9}$$

Table 8: Expected payoffs of messages contingent on the unknown state of the world

(a) Majority decisions

		Mess R	Mess B
40-10	State R	46.01	33.51
	State B	14.89	24.75
35-15	State R	46.72	37.06
	State B	15.94	20.88

(b) Unanimity decisions

		Mess R	Mess B
40-10	State R	46.04	27.54
	State B	15.38	27.32
35-15	State R	46.70	32.24
	State B	17.54	24.62

Note: Conditional continuation payoffs for sending message R or B, conditional on the true state.

and estimate (λ, α_m) by maximum likelihood using (6)–(9). Under correct Bayesian updating with prior $1/2$ and signal accuracy 0.6, the implied value is $\alpha_m \approx 0.81$, yielding posterior 0.6 after R and 0.4 after B. Overreaction corresponds to $\alpha_m > 0.81$.

Table 9 reports the resulting comparison between observed messaging frequencies and model-implied predictions without and with overreaction, as well as the implied beliefs $\Pr(\omega = R \mid s)$. The estimated α_m is around 5, far above the Bayesian benchmark. Interpreted literally, message behavior is consistent with subjects acting as if the state matches their private signal with probability above 0.9. This degree of perceived precision generates message predictions close to the observed frequencies.

Table 9: Predicted messages with and without overreaction

	Empirical	Beliefs under expected payoffs			Beliefs with overreaction		
	$\hat{\Pr}(m = R\|s)$	λ	$\Pr(\omega = R\|s)$	$\Pr(m = R\|s)$	(λ, α_m)	$\Pr(\omega = R\|s)$	$\Pr(m = R\|s)$
Majority (40-10 and 35-15 pooled)							
$s = R$	0.847	0.41	0.6	0.821	(0.20, 4.94)	0.92	0.877
$s = B$	0.146		0.4	0.495		0.08	0.233
Unanimity (40-10 and 35-15 pooled)							
$s = R$	0.949	0.33	0.6	0.879	(0.21, 5.57)	0.94	0.956
$s = B$	0.138		0.4	0.574		0.06	0.169

At the same time, message data alone do not sharply distinguish overreaction from lying aversion: both can rationalize near-truthful communication. The key difference is that overreaction also restricts behavior in the *voting* stage, whereas a pure honesty preference does not. We therefore turn next to voting behavior conditional on realized message profiles.

Result 2. *Messages are close to truthful and only weakly responsive to continuation-payoff incentives. Lying aversion and overreaction are both compatible with the observed messages in isolation.*

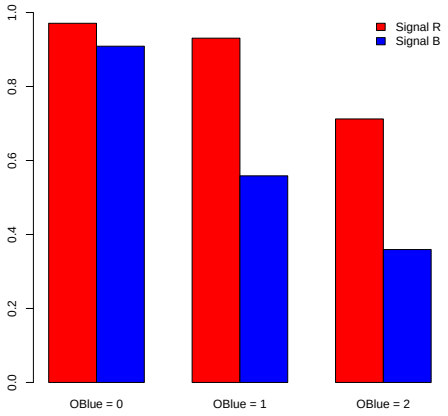
3.3 Do subjects vote rationally?

We now turn to the voting stage. Figure 1 summarizes voting behavior across information sets, conditioning on the subject's private signal $s_i \in \{R, B\}$ and on the number of *opponents'* messages equal to B, denoted $O \in \{0, 1, 2\}$. In both Majority and Unanimity treatments, voting responds strongly, and in the intuitive direction, to public information: subjects vote R with very high probability when

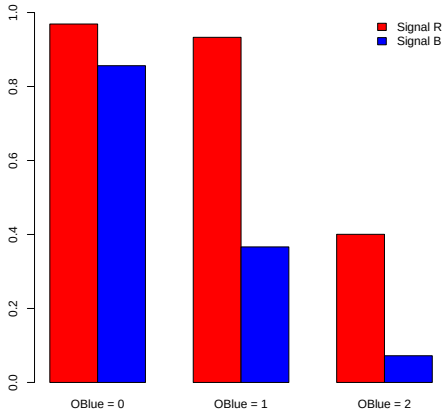
both opponents message R, with very low probability when both opponents message B, and at intermediate rates otherwise. Across information sets, the relative frequency of voting R ranges roughly from 0.1 to 0.95.

Figure 1: Relative frequency of voting R as functions of own signal and the number O of B-messages sent by opponents

(a) Majority treatments



(b) Unanimity treatments



The central question is whether this responsiveness is consistent with payoff maximization under rational expectations, given the expressive incentive to vote R. The answer hinges on two objects: (i) beliefs about the state conditional on the available information and (ii) pivotality, i.e. how often a vote changes the committee outcome. In particular, conditional on messages being highly informative (Section 3.2), the expressive incentive predicts substantial strategic R voting whenever pivotality is low. Conversely, if subjects overweight the informational content of signals and messages, they will be willing to forgo the expressive payoff more often and vote with message majorities even in low-pivotality situations.

An expected-payoff benchmark under rational expectations

To make the comparison operational without imposing equilibrium restrictions, we construct an empirical rational-expectations benchmark for each voting information set I (given by the subject's private signal and the observed message profile). The construction proceeds in two steps.

Step 1: beliefs about the state given information. Let $p(I) = \Pr(\omega = R \mid I)$ and $q(I) = 1 - p(I) = \Pr(\omega = B \mid I)$. We estimate $p(I)$ from realized states in the data using a flexible logit specification in the information available at the voting stage. Writing $O = \sum_{j \neq i} \mathbf{1}\{m_j = B\}$ for the number of opponents' B-messages, we use

$$p(I) \;=\; \frac{1}{1 + \exp\{\alpha_0 + \alpha_1 \cdot (\mathbf{1}\{s_i = B\} - 1/2) + \alpha_2 \cdot (O - 1)\}}. \tag{10}$$

This provides a compact reduced-form summary of how private signals and opponents' message

realizations predict the true state under empirically observed communication.

Step 2: outcome sensitivity to an individual's vote. Let $X \in \{R, B\}$ denote the committee outcome and let $M = \sum_{j=1}^{3} \mathbf{1}\{m_j = B\}$ denote the total number of B messages in the committee. We estimate the probability that the committee outcome is R conditional on the subject's vote v and the message profile (summarized by M):

$$\Pr(X = R \mid v, M) \;=\; \frac{1}{1 + \exp\{\beta_1 \cdot \mathbf{1}\{v = B\} + \sum_{k=0}^{3} \beta_{2|k} \cdot \mathbf{1}\{M = k\}\}}. \tag{11}$$

Allowing the intercept to vary flexibly with M captures the empirically relevant dependence of outcomes on message profiles and accommodates noise in the voting stage induced by opponents' stochastic voting.

Given (10)–(11), the probability that the committee decision is correct conditional on voting v is

$$\Pr(X = \omega \mid v, I) = p(I) \Pr(X = R \mid v, M) + q(I) \Pr(X = B \mid v, M), \tag{12}$$

with $\Pr(X = B \mid v, M) = 1 - \Pr(X = R \mid v, M)$. Using the payoff parameters (C, K) from Section 2.1, expected payoffs are

$$EP(v = R \mid I) = C \Pr(X = \omega \mid v = R, I) + K, \qquad EP(v = B \mid I) = C \Pr(X = \omega \mid v = B, I),$$

under Majority. Under Unanimity with default-R enforcement, the expressive payoff is received if and only if the enforced final vote is R, which coincides with $X = R$ under our reduced-form implementation. Hence we use

$$EP(v \mid I) = C \Pr(X = \omega \mid v, I) + K \Pr(X = R \mid v, M). \tag{13}$$

Table 10 juxtaposes, for each voting information set, the expected-payoff benchmark under rational expectations with the corresponding vote frequencies in the data. Two features are robust.

First, under Majority, the expected-payoff benchmark typically favors R. A key reason is that the expressive payoff K is attached to the *individual vote* under Majority: if a subject votes R she receives K regardless of whether her vote is pivotal. When pivotality is low, the sure expressive payoff dominates the expected common-value gain from trying to shift the outcome from R to B, so $EP(R) > EP(B)$ in most information sets.

Second, subjects nevertheless vote B precisely when the information favors B. Even when $EP(R) > EP(B)$ under the rational-expectations benchmark, B votes are frequent in the information sets in which the message profile (and often also the subject's signal) indicates that $\omega = B$ is likely. Under Unanimity, incentives to vote B differ because the expressive payoff is tied to the enforced final vote (equivalently, to the default-R outcome) rather than to the raw individual vote. Nevertheless, the same qualitative pattern remains: subjects sometimes vote B in information sets where the expected-payoff benchmark favors R. In short, votes track signals and messages in the right direction, but they do so *more aggressively* than payoff maximization under rational expectations would justify.

Table 10: Expected payoffs and decisions when voting

	Majority 40-10			Majority 35-15			Unanimity 40-10			Unanimity 35-15		
	$EP(R)$	$EP(B)$	$\hat{\Pr}(R)$	$EP(R)$	$EP(B)$	$\hat{\Pr}(R)$	$EP(R)$	$EP(B)$	$\hat{\Pr}(R)$	$EP(R)$	$EP(B)$	$\hat{\Pr}(R)$
S-R M-R , OBlue 0	38.17 (0.57)	26.4 (0.49)	0.98 (0.007)	39.44 (0.54)	23.38 (0.48)	0.98 (0.008)	39.17 (0.55)	37.21 (0.49)	0.98 (0.007)	40.74 (0.5)	39.56 (0.48)	0.99 (0.004)
S-R M-R , OBlue 1	33.67 (0.38)	22.64 (0.29)	0.93 (0.011)	35.65 (0.34)	19.89 (0.27)	0.95 (0.009)	33.73 (0.4)	31.82 (0.33)	0.95 (0.009)	35.6 (0.36)	33.45 (0.32)	0.96 (0.009)
S-R M-R , OBlue 2	29.61 (0.22)	20.89 (0.49)	0.64 (0.031)	31.96 (0.38)	17.94 (0.3)	0.75 (0.028)	25.71 (0.21)	23 (0.56)	**0.3** (**0.033**)	27.32 (0.28)	22.29 (0.33)	**0.48** (**0.031**)
S-R M-B , OBlue 0	38.17 (0.57)	26.4 (0.49)	0.86 (0.056)	39.44 (0.54)	23.38 (0.48)	0.91 (0.04)	39.17 (0.55)	37.21 (0.49)	0.71 (0.113)	40.74 (0.5)	39.56 (0.48)	**0.47** (**0.118**)
S-R M-B , OBlue 1	32.35 (0.27)	18.29 (0.21)	0.68 (0.058)	35.18 (0.29)	17.47 (0.12)	0.94 (0.023)	29.72 (0.27)	20.52 (0.29)	0.53 (0.109)	32.75 (0.3)	23.47 (0.28)	0.62 (0.061)
S-R M-B , OBlue 2	29.61 (0.22)	20.89 (0.49)	0.55 (0.111)	31.96 (0.38)	17.94 (0.3)	0.84 (0.039)	25.71 (0.21)	23 (0.56)	**0.25** (**0.125**)	27.32 (0.28)	22.29 (0.33)	0.54 (0.144)
S-B M-R , OBlue 0	31.21 (0.67)	20.95 (0.52)	0.77 (0.057)	33.41 (0.63)	18.27 (0.53)	0.92 (0.057)	32.03 (0.77)	31.06 (0.67)	0.74 (0.062)	34.54 (0.6)	33.83 (0.57)	0.92 (0.032)
S-B M-R , OBlue 1	26.33 (0.38)	17.36 (0.29)	0.82 (0.037)	29.35 (0.34)	15.11 (0.27)	0.78 (0.05)	25.99 (0.4)	25.78 (0.31)	0.87 (0.039)	28.91 (0.36)	28.11 (0.3)	0.81 (0.049)
S-B M-R , OBlue 2	27.4 (0.35)	25.96 (0.45)	0.51 (0.073)	28.35 (0.39)	20.83 (0.33)	0.69 (0.06)	23.89 (0.16)	28.11 (0.41)	0.31 (0.079)	24.6 (0.22)	25.47 (0.25)	0.31 (0.071)
S-B M-B , OBlue 0	31.21 (0.67)	20.95 (0.52)	0.92 (0.015)	33.41 (0.63)	18.27 (0.53)	0.92 (0.019)	32.03 (0.77)	31.06 (0.67)	0.87 (0.02)	34.54 (0.6)	33.83 (0.57)	0.84 (0.022)
S-B M-B , OBlue 1	27.65 (0.27)	21.71 (0.21)	**0.44** (**0.021**)	29.82 (0.29)	17.53 (0.12)	0.6 (0.021)	25.54 (0.21)	24.49 (0.2)	**0.2** (**0.017**)	27.86 (0.27)	24.44 (0.11)	**0.38** (**0.019**)
S-B M-B , OBlue 2	27.4 (0.35)	25.96 (0.45)	**0.31** (**0.03**)	28.35 (0.39)	20.83 (0.33)	**0.31** (**0.026**)	23.89 (0.16)	28.11 (0.41)	0.03 (0.01)	24.6 (0.22)	25.47 (0.25)	0.05 (0.013)

Note: For each voting information set, the table reports the expected payoff from voting R, $EP(R)$, the expected payoff from voting B, $EP(B)$, and the observed relative frequency of R votes, $\hat{\Pr}(R)$. Expected payoffs are computed from the empirical belief and outcome objects estimated in (10)–(13).

Result 3. *Voting is highly responsive to signals and message profiles, but it is not well-explained by payoff maximization under rational expectations: subjects vote for B too often relative to the expected-payoff benchmark, especially when public information favors B.*

A simple diagnostic: what beliefs are required to justify B-voting?

The comparison above is informative but does not by itself isolate *what* departs from rational voting: subjects may misperceive pivotality, misperceive the informativeness of signals/messages, or both. Voting against the expressive option R is costly unless the subject both (i) assigns sufficiently high probability to $\omega = B$ and (ii) expects her vote to matter for the outcome.

A convenient way to summarize outcome sensitivity from (11) is the effect of switching one's vote on the probability of outcome R:

$$\pi(I) \;\equiv\; \Pr(X = R \mid v = R, M) - \Pr(X = R \mid v = B, M). \tag{14}$$

Under Majority, $\pi(I)$ is small whenever the other votes typically determine the outcome regardless of i's vote. Under Unanimity with default-R enforcement, $\Pr(X = R \mid v = R, M)$ is mechanically close to one, so $\pi(I)$ is essentially the probability that voting R prevents a unanimous B outcome.

Using (12), one can write the payoff gain from voting B rather than R under Majority as

$$EP(B) - EP(R) \;=\; \pi(I) \cdot C \cdot \big(2q(I) - 1\big) \;-\; K. \tag{15}$$

Therefore voting B can be optimal under Majority only if the posterior $q(I)$ satisfies the threshold

$$q(I) \geq q^{\min}(I) \equiv \frac{1}{2}\Big(1+\frac{K}{\pi(I)C}\Big), \tag{16}$$

with the convention that if $\pi(I)=0$ then voting B is never optimal. The key implication is immediate: when pivotality is low, $q^{\min}(I)$ becomes extremely close to one (and may exceed one), so observing substantial B voting in such information sets requires *very extreme* beliefs that $\omega = B$.

Under Unanimity with enforcement, the expressive term in (13) is also scaled by the probability that the vote affects the outcome, so both the common-value gain and the expressive cost are proportional to $\pi(I)$. As a result, the corresponding posterior threshold is much less sensitive to pivotality (indeed, under the mechanical benchmark it reduces to a constant cutoff in $q(I)$). This difference is one reason why Unanimity can generate more B voting than Majority even when outcome sensitivity is low.

These implied-belief considerations foreshadow the structural analysis in Section 4. The message data indicate that subjects treat messages as informative; the voting data indicate that, given the expressive incentive, subjects behave as if the evidence about ω were more decisive (or outcome sensitivity higher) than the rational-expectations benchmark suggests. The structural section puts this joint discipline into a single likelihood framework and asks which deviations from the Bayesian benchmark are needed once messages and votes are fitted together.

4 Why do subjects deviate from rational behavior?

The empirical patterns documented above pose a *joint* challenge. Committee decisions are "too accurate" relative to Bayesian equilibrium with expressive payoffs, and this excess accuracy is generated in *both* stages: (i) subjects communicate more truthfully than is justified by expected-payoff incentives, and (ii) they vote more "sincerely" (i.e. more in line with signals and message profiles) than would maximize their own expected payoffs given the same information and the expressive incentive to vote R.

These two facts already narrow the menu of explanations. Lying aversion can rationalize truth-telling in the message stage, but it leaves the voting-stage pivotality/free-riding logic intact and therefore cannot, by itself, generate the observed responsiveness of votes to message profiles. Conversely, a shallow-reasoning account can inflate perceived pivotality and thereby push votes toward sincerity, but it would typically predict weaker use of others' messages, contradicting the strong empirical sensitivity of votes to public information. A belief distortion of the kind formalized in Section 2.2 ("overreaction") naturally moves both stages in the right direction: if agents overweight directional evidence about the state, then sending an accurate message becomes more valuable and voting with the informational majority becomes more attractive.

Accordingly, we treat the structural exercise as *disciplined measurement*. We embed a parsimonious belief distortion into an otherwise standard logit-response model of messaging and voting,

estimate it jointly, and use likelihood-ratio restrictions to assess which departures from the Bayesian benchmark are required by the data. The main estimates and restriction tests are summarized in Tables 11 and 12.

4.1 A structural model of messaging and voting

We merge the belief objects used above for (i) beliefs about the true state and (ii) beliefs about the mapping from votes to outcomes, and model choices in both stages via logit response. The unifying idea is that the same distortion that makes agents *too responsive* to evidence about the state can affect incentives in *both* stages.

Voting. Fix a voting information set I, consisting of the subject's private signal and the observed message profile. Let $\Pr(\omega = R \mid I)$ denote the subject's belief about the state and let $\Pr(X = R \mid v, I)$ denote the subject's belief about the committee outcome conditional on her own vote $v \in \{R, B\}$. Given these objects, define the perceived probability of a correct outcome under vote v by

$$\Pr(X = \omega \mid v, I) \;=\; \Pr(\omega = R \mid I)\Pr(X = R \mid v, I) + \Pr(\omega = B \mid I)\Pr(X = B \mid v, I),$$

with $\Pr(X = B \mid v, I) = 1 - \Pr(X = R \mid v, I)$. Under Majority, the expressive payoff is tied to the *raw individual vote*, so perceived expected payoffs are

$$EP(v = R \mid I) = C\Pr(X = \omega \mid v = R, I) + K, \qquad EP(v = B \mid I) = C\Pr(X = \omega \mid v = B, I).$$

Under Unanimity with default-R enforcement, the expressive payoff is received if and only if the enforced final vote is R, which coincides with $X = R$ under our reduced-form implementation. Hence,

$$EP(v \mid I) = C\Pr(X = \omega \mid v, I) + K\Pr(X = R \mid v, I).$$

To keep the structural interpretation transparent, we allow the data to scale the common-value and expressive components differently in the voting stage. Let

$$\Delta_C(I) \;\equiv\; C\big[\Pr(X = \omega \mid v = R, I) - \Pr(X = \omega \mid v = B, I)\big],$$

and let

$$\Delta_K(I) \equiv \begin{cases} K, & \text{under Majority,} \\ K\big[\Pr(X = R \mid v = R, I) - \Pr(X = R \mid v = B, I)\big], & \text{under Unanimity with default-}R\text{ enforcement.} \end{cases}$$

We then specify the probability of voting R at I as

$$\Pr(v = R \mid I) \;=\; \frac{1}{1 + \exp\big\{-\lambda_v \cdot \big(\Delta_C(I) + \phi_v \Delta_K(I)\big)\big\}}, \tag{17}$$

where $\lambda_v > 0$ is a payoff-sensitivity parameter and ϕ_v governs the relative weight placed on the

expressive component. The restriction $\phi_v = 1$ corresponds to correctly weighting the expressive payoff as specified in (2); $\phi_v = 0$ corresponds to purely accuracy-motivated voting, holding beliefs fixed.

Beliefs about the state. Write $O = \sum_{j \neq i} \mathbf{1}\{m_j = B\} \in \{0,1,2\}$ for the number of opponents' B-messages. We parameterize state beliefs by a logit index in the two sufficient statistics observed at the voting stage, s_i and O:

$$\Pr(\omega = R \mid I) \;=\; \frac{1}{1 + \exp\{\delta_0 + \delta_s \cdot (\mathbf{1}\{s_i = B\} - 1/2) + \delta_m \cdot (O - 1)\}}. \tag{18}$$

This specification is deliberately agnostic about equilibrium restrictions; it provides a low-dimensional summary of how the subject treats private and public evidence about the state. The belief-distortion restriction below imposes structure on (δ_s, δ_m).

Beliefs about the outcome. Let $M = \sum_{j=1}^{3} \mathbf{1}\{m_j = B\} \in \{0,1,2,3\}$ denote the total number of B messages in the committee (including the subject). We parameterize beliefs about the outcome mapping by allowing the baseline propensity for outcome R to vary flexibly with M, and allowing the subject's own vote to shift this propensity:

$$\Pr(X = R \mid v, I) \;=\; \frac{1}{1 + \exp\{\beta_1 \cdot \mathbf{1}\{v = B\} + \sum_{k=0}^{3} \beta_{2|k} \cdot \mathbf{1}\{M = k\}\}}. \tag{19}$$

This reduced-form object captures the empirically relevant mapping from individual votes to outcomes given the message profile, allowing for stochastic voting by opponents.

Messaging. Let $dEP(s)$ denote the perceived expected payoff difference between sending $m = B$ and $m = R$ after observing signal s, computed as in (7)–(8) using the subject's state belief after the signal. We model message choice via

$$\Pr(m = R \mid s) \;=\; \frac{1}{1 + \exp\big\{\lambda_m \cdot dEP(s) - \eta \cdot \big(\mathbf{1}\{s = R\} - \mathbf{1}\{s = B\}\big)\big\}}, \tag{20}$$

where λ_m captures responsiveness to instrumental continuation-payoff incentives and η captures a direct truth-telling motive (our "lying-aversion" channel).

Interpreting overreaction as perceived precision. The belief distortion studied in Section 2.2 scales posterior log-odds, equivalently raising likelihood ratios to a common power. In the present parametrization, this corresponds to a simple restriction on the *relative* weights placed on private and public evidence in (18). Let $(\hat{\delta}_s, \hat{\delta}_m)$ denote the Bayesian benchmark weights implied by Bayes' rule given the objective signal structure and the empirically relevant (symmetric) communication

technology.[7] We impose

$$(\delta_s, \delta_m) \;=\; \kappa \cdot (\hat{\delta}_s, \hat{\delta}_m), \tag{21}$$

where $\kappa = 1$ corresponds to Bayesian beliefs, $\kappa > 1$ corresponds to "overreaction" (inflated perceived precision), and $\kappa \in (0,1)$ corresponds to underreaction. Importantly, (21) does *not* impose that subjects ignore others' information; it imposes that they scale *all* directional evidence by the same factor, consistent with the log-odds scaling property shown in Appendix A.4.

4.2 Estimation, pooling, and experience splits

All parameters are estimated by maximum likelihood. Standard errors are Huber–Sandwich estimates clustered at the subject level. Hypothesis tests for economically interpretable restrictions are likelihood-ratio tests, with critical values obtained from a subject-level block bootstrap to account for within-subject dependence across rounds. Table 11 reports both parameter estimates and restriction tests.

We begin by estimating each treatment in isolation and testing whether the two payoff calibrations can be pooled within each voting rule. The data do not reject pooling within Majority and within Unanimity (all $p > 0.8$), so we pool within rule while controlling for the payoff parameter K. Because each pooled rule cell contains 93 subjects and 50 games per subject, splitting the sample into the first and second halves yields $93 \times 25 = 2325$ subject–round observations per rule and half-session. Across Majority/Unanimity and first/second halves, this produces $4 \times 2325 = 9300$ observations in the four reported subsamples.

We then test for experience effects. While within-treatment learning is only modestly detectable, pooling within rule reveals statistically significant differences between the first and second halves of sessions (Majority: $p = 0.027$; Unanimity: $p = 0.035$). We therefore report estimates separately for periods 1–25 and 26–50 throughout.

4.3 Analysis

The restriction tests mirror the reduced-form narrative. We proceed in three steps: we first test whether choices can be rationalized by reweighting payoff components, then we test whether beliefs about the state satisfy the proportional "precision distortion" in (21), and finally we study whether beliefs about the mapping from messages and votes to outcomes display a pivotality-misperception pattern.

Payoff weights. Row (A) of Table 11 tests whether subjects disproportionately weight the expressive component relative to the common-value component in the voting stage (i.e. $\phi_v \neq 1$ in (17)) and whether payoff sensitivity differs across voting and messaging (i.e. $\lambda_v \neq \lambda_m$). The restriction is not rejected in any subsample (all $p > 0.2$). Thus, the data do not call for "efficiency concerns"

[7] Concretely, $\hat{\delta}_s$ equals the log-likelihood ratio induced by one private signal, and $\hat{\delta}_m$ equals the log-likelihood ratio induced by one opponent message, computed from the conditional distribution of messages given the state.

Table 11: Structural estimation: motives underlying communication and voting

		Majority decision				Unanimous decision			
		First half		Second half		First half		Second half	
Null Hypothesis	Alternative	Rational	Behav	Rational	Behav	Rational	Behav	Rational	Behav
α_m		0.771 (0.086)	22.601 (181.016)	0.746 (0.086)	3.388 (4.223)	0.738 (0.086)	25.379 (384.213)	0.875 (0.088)	4.919 (3.146)
α_1		0.574 (0.074)	12.248 (90.552)	0.556 (0.074)	−0.515 (1.266)	0.449 (0.081)	14.511 (192.381)	0.893 (0.08)	4.201 (2.169)
α_2		0.377 (0.074)	12.021 (90.352)	0.469 (0.075)	4.151 (3.853)	0.574 (0.067)	13.559 (192.428)	0.644 (0.074)	3.68 (1.735)
β_1		2.407 (0.094)	4.746 (0.475)	2.397 (0.101)	3.092 (0.373)	2.236 (0.103)	5.105 (2.018)	2.169 (0.1)	6.573 (3.542)
$\beta_{2\|0}$		−3.466 (0.189)	−1.837 (0.688)	−4.361 (0.297)	0.083 (0.09)	−3.98 (0.183)	2.731 (1.282)	−4.323 (0.265)	4.701 (1.987)
$\beta_{2\|1}$		−2.678 (0.091)	−3.406 (0.427)	−3.434 (0.115)	0.299 (0.399)	−3.23 (0.112)	3.471 (1.295)	−3.352 (0.116)	5.137 (1.996)
$\beta_{2\|2}$		−0.387 (0.059)	−2.795 (0.158)	−0.791 (0.062)	−3.818 (1.173)	−0.247 (0.062)	−3.709 (1.458)	−0.513 (0.059)	5.02 (2.146)
$\beta_{2\|3}$		0.861 (0.096)	−1.8 (0.338)	0.394 (0.094)	1.425 (1.216)	0.768 (0.112)	−0.243 (16.197)	1.104 (0.104)	0.216 (41.164)
λ_1			0.969 (0.063)		1.218 (0.275)		1.542 (0.257)		3.63 (0.768)
λ_2			1.105 (0.164)		0.38 (0.183)		2.007 (0.27)		6.341 (1.47)
λ_3			0.029 (1.034)		0 (1.055)		1.501 (0.403)		1.778 (0.847)
λ_4			1.668 (1.055)		1.774 (0.769)		0.244 (0.505)		0.663 (0.915)
No of observations			2325		2325		2325		2325
Log-likelihood			-2081.1		-1905.06		-1632.48		-1516.48
Efficiency concerns									
(A) $H_A : \lambda_1 = \lambda_2 = \lambda_3$	H_{Base}		3.05 (0.605)		9.35 (0.28)		0.83 (0.795)		6.84 (0.244)
State beliefs: overshooting									
(B) $H_B : H_A \wedge \alpha_{m,1,2} = \hat{\alpha}_{m,1,2}$	H_A		37.62*** (0.006)		39.79** (0.033)		98.44*** (0.001)		50.29** (0.042)
(C) $H_C : H_A \wedge \alpha_{m,1,2} \propto \hat{\alpha}_{m,1,2}$	H_A		4.46 (0.258)		26.26 (0.143)		8.39* (0.07)		6.25 (0.44)
Voting beliefs: rational expectations with pivotality illusion									
(D) $H_D : H_C \wedge \beta_1 = \hat{\beta}_1$	H_C		0 (1)		0 (1)		4.03 (0.442)		11.7 (0.106)
(E) $H_E : H_D \wedge \beta_{2\|0\ldots3} = \hat{\beta}_{2\|0\ldots3}$	H_D		32.69*** (0)		30.44*** (0)		340.34*** (0)		301.08*** (0)
(F) $H_F : H_D \wedge \beta_{2\|0\ldots3} \propto \hat{\beta}_{2\|0\ldots3}$	H_D		15.09** (0.019)		9.35* (0.088)		31.66** (0.036)		8.94 (0.3)
(G) $H_G : H_D \wedge \beta_{2\|0\ldots3} = 0$	H_D		16.46** (0.032)		10.78* (0.092)		31.66** (0.036)		10 (0.27)
Lying aversion develops with experience									
(H) $H_H : H_D \wedge \lambda_4 = 0$	H_D		19.91 (0.109)		37.94* (0.058)		0.1 (0.916)		17.55 (0.194)

Note: The rows α_m–λ_4 report maximum-likelihood estimates; standard errors in parentheses are Huber–White estimates clustered at the subject level. The columns labelled "Rational" report the benchmark coefficients implied by rational expectations and Bayesian updating (using $\hat{\alpha}$ and $\hat{\beta}$ as defined in the text), while the columns labelled "Behav" report the corresponding estimates in the flexible specification. Rows (A)–(H) report likelihood-ratio (LR) tests of nested restrictions. The reported statistic is the LR statistic $2(\ell_{\text{unrestricted}} - \ell_{\text{restricted}})$; p-values (in parentheses) are obtained by a subject-level bootstrap of the LR statistic to account for within-subject dependence across rounds. Asterisks denote significance levels (* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$). The notation $\alpha_{m,1,2}$ denotes the vector $(\alpha_m, \alpha_1, \alpha_2)$ and $\propto$ denotes proportionality of vectors.

(e.g. $\phi_v < 1$) or for disproportionate salience of K (e.g. $\phi_v > 1$) as the primary driver. The structural explanation therefore points to beliefs rather than to payoff weights.

State beliefs: Bayesian weights versus perceived precision. Row (B) tests Bayesian state beliefs by imposing $(\delta_s, \delta_m) = (\hat{\delta}_s, \hat{\delta}_m)$ in (18). This restriction is rejected across subsamples (reported p-values range from 0.001 to 0.042). Substantively, the estimated evidence weights are much larger in magnitude than their Bayesian counterparts, consistent with agents behaving as if both private signals and messages are substantially more precise than under the objective information structure.

Relative weighting of own signal and opponents' messages. A cursedness-/level-k-style deviation would predict that opponents' messages are treated as *less* informative than they objectively are, i.e. the relative weight on one's own signal should rise compared to the Bayesian benchmark. Row (C) addresses this by testing the proportionality restriction in (21), which preserves the *ratio* of message- to signal-weight while allowing a common amplification factor κ. The proportionality restriction is not rejected in most subsamples; where it is weakly rejected, the direction does not support systematic message discounting. The data therefore do not look "cursed" in the sense of selectively underweighting others' information. Rather, they are consistent with an approximately common scaling of *all* directional evidence, i.e. an intermediate $\kappa > 1$.

The remaining rows turn from beliefs about the state to beliefs about the *committee outcome* induced by voting.

Perceived impact of one's own vote. Row (D) tests whether subjects correctly perceive the sign and magnitude of the effect of their own vote on the probability of outcome R by imposing $\beta_1 = \hat{\beta}_1$ in (19). The restriction is not rejected. This is useful because it rules out a simple misunderstanding of the voting rule as the main explanation for "sincere" voting.

Rational expectations about how messages translate into outcomes. Row (E) tests rational expectations about the outcome mapping conditional on message profiles by imposing $\beta_{2|\cdot} = \hat{\beta}_{2|\cdot}$. This restriction is strongly rejected. Thus, even if subjects are roughly correct about how their *own* vote shifts the outcome, they mispredict how others' votes respond to messages and, therefore, how message profiles translate into final outcomes.

Structure of the outcome-belief deviation: a pivotality illusion. Rows (F) and (G) probe the structure of this misprediction. With experience, subjects' outcome-belief weights are difficult to distinguish from either (i) a proportional attenuation of the rational-expectations weights, $\beta_{2|\cdot} = \rho\,\hat{\beta}_{2|\cdot}$ with $\rho \in [0,1]$, or (ii) the extreme simplification $\beta_{2|\cdot} = 0$, in which message profiles are treated as largely uninformative about others' votes and hence about outcomes. Both patterns point in the same qualitative direction: subjects behave as if others' votes are noisier and less predictable from messages than they truly are, which inflates perceived pivotality and thereby encourages voting in line

with informational considerations. Importantly, this is not a level-k story about the *message* stage (which would also dismiss message informativeness); it is better described as a *pivotality illusion* localized to the mapping from messages to outcomes.

Is lying aversion required once belief distortions are allowed? Row (H) tests whether a direct truth-telling motive is needed in the messaging stage by setting $\eta = 0$ in (20) once state-belief distortions are admitted. The restriction is not robustly rejected. Any weak detection in isolated subsamples is not stable across specifications and does not affect the substantive conclusion: the joint pattern of (near-truthful) messages and (strongly responsive) votes is primarily accounted for by belief distortions and outcome-belief misprediction, not by a direct utility cost of lying.

Result 4. *The structural restrictions align with the reduced-form evidence. Subjects behave as if they (i)* overreact *when forming beliefs about the state, in the sense of an amplified perceived precision* $\kappa > 1$ *applied broadly to both signals and messages, and (ii)* under-infer *how predictably message profiles translate into the committee outcome (a pivotality-illusion channel), especially with experience. Once (i) is admitted, lying aversion is not a robust additional driver.*

4.4 Quantitative fit

Table 12: Adequacy of simple models in describing behavior (Majority and Unanimity pooled)

	1st halves of sessions					2nd halves of sessions				
	LL	R^2	γ_1	γ_2	γ_3	LL	R^2	γ_1	γ_2	γ_3
QRE	-5546.63	0.35	1	0	1	-5045.16	0.47	1	0	1
+ overreaction	-4123.14	0.87	6.15	0	1	-3860.76	0.87	3.97	0	1
+ lying aversion	-4573.18	0.71	1	1.36	1	-4175.56	0.77	1	0.90	1
+ pivotality illusion	-5102.90	0.49	1	0	0	-4715.25	0.57	1	0	0.19

Note: The table compares nested parsimonious specifications fit to the joint distribution of messages and votes, pooling Majority and Unanimity within each experience half. γ_1 denotes the perceived-precision distortion κ (equivalently the common ratio of estimated to Bayesian evidence weights, e.g. $\delta_s/\hat{\delta}_s = \delta_m/\hat{\delta}_m$ under (21)). γ_2 indexes the strength of a direct truth-telling motive in messaging (a normalized version of η in (20)). γ_3 indexes attenuation of outcome-belief responsiveness to message profiles (a normalized version of ρ in $\beta_{2|\cdot} = \rho\,\hat{\beta}_{2|\cdot}$), with smaller values corresponding to a stronger pivotality-illusion channel.

Table 12 compares four nested parsimonious specifications (baseline logit-QRE; plus overreaction; plus lying aversion; plus pivotality illusion) separately for the first and second halves of sessions. Two patterns are immediate. First, allowing the single perceived-precision distortion produces by far the largest improvement in fit in both halves. Second, the ranking is stable across experience: overreaction dominates, lying aversion improves fit less, and the pivotality-illusion channel by itself captures comparatively little of the joint variation.

These likelihood comparisons translate directly into standard information criteria. Let LL denote the maximized log-likelihood, k the number of free parameters, and n the number of observations. Then $\text{AIC} = 2k - 2LL$ and $\text{BIC} = k\ln(n) - 2LL$. In each half, pooling Majority and Unanimity yields $n = 4650$ observations; pooling across both halves yields $n = 9300$. Since the overreaction

specification improves *LL* by more than a thousand relative to baseline in each half (Table 12), no plausible AIC/BIC penalty for adding a single parameter can overturn the ranking.[8]

Result 5. *Measured by likelihood (and therefore also by AIC/BIC), the dominant parsimonious improvement over baseline logit-QRE is the one-parameter perceived-precision (overreaction) channel. This conclusion is stable across early and late rounds.*

Finally, this section does not claim that *every* deviation is captured by a single distortion. Rather, the disciplined takeaway is that a single perceived-precision distortion accounts for the main joint anomaly across stages, while remaining deviations are naturally summarized by a localized pivotality-illusion component in beliefs about how message profiles translate into outcomes.

5 Discussion

This paper documents a robust deviation from Bayesian equilibrium in a standard committee game with cheap-talk communication and expressive voting incentives. Committees aggregate private information substantially more accurately than Bayesian equilibrium predicts under either Majority or Unanimity. The data show two stage-by-stage regularities. First, messages are close to truthful (Section 3.2). Second, conditional on message profiles, votes respond strongly, and often more strongly than payoff maximization under rational expectations would imply, to both private and public information (Section 3.3). The key empirical discipline comes from the voting stage: the observed responsiveness of votes implies posteriors that are systematically more extreme than the Bayesian benchmark, in the direction of overweighting new information.

To organize these patterns parsimoniously, we estimate a structural model that layers a one-parameter distortion of Bayesian updating onto a standard logit-response framework for messaging and voting (Section 4). The distortion can be read as an index of responsiveness to directional evidence, or equivalently as an "as-if" perceived precision of signals and messages. Allowing this single belief-distortion parameter materially improves fit relative to a Bayesian benchmark and, crucially, rationalizes behavior in *both* stages jointly. Additional components, a localized misprediction of how message profiles map into outcomes (a pivotality-illusion channel) and, at most, weak direct truth-telling motives, matter at the margin, but the central empirical regularity is that a common amplification of evidence about the state provides the main unifying account of (i) near-truthful communication and (ii) overly responsive voting.

What the design can and cannot establish. Our contribution is measurement and organization, not a clean causal separation among all candidate models of non-Bayesian belief formation. The design was not constructed to toggle individual belief mechanisms in isolation. Accordingly, the structural exercise should be read as a disciplined accounting within a canonical committee environment: it identifies the direction and magnitude of systematic departures from Bayesian updating

[8] The penalty difference between two specifications that differ by one parameter is 2 under AIC and $\ln(n)$ under BIC. Here $\ln(4650) \approx 8.44$ for a half-sample and $\ln(9300) \approx 9.14$ when pooling across halves.

that are needed to satisfy the *joint* restrictions imposed by messages and votes. Other psychological mechanisms may generate similar reduced-form distortions, and a different design would be required to adjudicate among them.

Alternative mechanisms. Two prominent alternatives are honesty motives and limited strategic reasoning. Honesty motives can rationalize truthful messaging, but by themselves they do not explain why votes behave as if posteriors are systematically more extreme than Bayes conditional on informative message profiles. Limited-depth reasoning can, in principle, distort perceived pivotality and push votes toward information use, but standard versions also predict weaker reliance on others' messages, which conflicts with the strong empirical sensitivity of votes to message profiles. Risk aversion is also unlikely to explain our findings: it reduces the effective weight on the risky common-value component relative to the sure expressive payoff, which directionally strengthens incentives to vote R and thereby works *against* information aggregation. Social preferences could increase truth-telling if subjects dislike harming others through deception, but they do not naturally generate the second-stage implication that voting reflects systematically inflated posteriors. These considerations motivate our focus on belief distortions as the most economical stage-consistent account.

Toward cleaner tests. A sharper causal test of belief distortions versus competing mechanisms would vary the informational environment in ways that shift Bayesian posteriors while holding strategic incentives fixed. Natural design elements include individual ($N = 1$) decision problems to isolate updating, exogenously biased priors that are displayed prominently to test base-rate neglect versus overreaction, and sender–receiver variants with exogenous message accuracy (or experimentally controlled truthfulness) to break equilibrium feedback from voting incentives to communication incentives. Such designs would complement the present results by turning the reduced-form "responsiveness" parameter into a cleaner discriminator among specific non-Bayesian mechanisms.

Portability. The belief-distortion interpretation is not specific to expressive payoffs. A recurring finding in experimental work on committee communication is "overcommunication" and heavy reliance on messages relative to equilibrium predictions. A parsimonious distortion of Bayesian updating provides a natural organizing language for these findings because it maps directly into exaggerated posteriors from private signals and, conditional on informative messages, exaggerated posteriors from message profiles. This perspective may be useful for future work on committee design, where comparative statics in decision rules and communication protocols depend critically on how committee members form beliefs from private and public information.

6 Conclusion

We study information aggregation in three-person committees with private signals, a cheap-talk communication stage, and voting under Majority or Unanimity rules, in an environment where expressive incentives create scope for both strategic communication and strategic voting. The experiment reveals

a robust discrepancy between Bayesian equilibrium predictions and observed behavior: committees are significantly more accurate than predicted, messages are close to truthful, and votes respond to signals and message profiles more strongly than payoff maximization under rational expectations would imply.

The paper's contribution is to document this "too accurate" committee behavior in a canonical laboratory environment with transparent incentives and rich observables, and to provide a parsimonious structural account that organizes behavior across both stages. A one-parameter perceived-precision distortion layered onto a logit-response model captures most of the systematic joint variation in messages and votes and implies posteriors that are systematically more extreme than Bayes in the direction of overweighting new information. Residual deviations are naturally summarized by a localized misprediction of how message profiles map into outcomes (a pivotality-illusion channel).

At the same time, the evidence should not be read as a definitive causal separation among all non-Bayesian updating models or among all potential moral and social motivations. Establishing sharper causal distinctions would require follow-up designs that exogenously vary priors and information, or that isolate belief updating from strategic feedback. Within the present committee environment, however, the disciplined conclusion is that belief distortions provide the most economical account of the joint restrictions implied by messaging and voting behavior.

7 Declarations

Funding

This study was funded by the WZB Berlin and the DFG (project BR 4648/1 and CRC TRR 190).

Competing Interests

The authors have no relevant financial or non-financial interests to disclose.

References

Afrouzi, H., Kwon, S. Y., Landier, A., Ma, Y., and Thesmar, D. (2023). Overreaction in Expectations: Evidence and Theory*. *The Quarterly Journal of Economics*, 138(3):1713–1764.

Ali, S. N., Goeree, J. K., Kartik, N., and Palfrey, T. R. (2008). Information aggregation in standing and ad hoc committees. *American Economic Review*, 98(2):181–86.

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

Bar-Hillel, M. (1980). The base-rate fallacy in probability judgments. *Acta Psychologica*, 44(3):211–233.

Bhattacharya, S. (2013). Preference monotonicity and information aggregation in elections. *Econometrica*, 81(3):1229–1247.

Bordalo, P., Gennaioli, N., Ma, Y., and Shleifer, A. (2020). Overreaction in macroeconomic expectations. *American Economic Review*, 110(9):2748–82.

Breitmoser, Y. and Valasek, J. (2024). Strategic communication in committees with expressive payoffs. *The RAND Journal of Economics*, 55(1):33–54.

Cai, H. and Wang, J. T.-Y. (2006). Overcommunication in strategic information transmission games. *Games and Economic Behavior*, 56(1):7–36.

Callander, S. (2007). Bandwagons and Momentum in Sequential Voting. *The Review of Economic Studies*, 74:653–684.

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

Charness, G. and Dufwenberg, M. (2006). Promises and partnership. *Econometrica*, 74(6):1579–1601.

Dal Bó, E. (2007). Bribing voters. *American Journal of Political Science*, 51(4):789–803.

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.

De Filippis, R., Guarino, A., Jehiel, P., and Kitagawa, T. (2022). Updating ambiguous beliefs in a social learning experiment. *Journal of Economic Theory*, 199:105371.

Ekmekci, M. and Lauermann, S. (2019). Manipulated electorates and information aggregation. *The Review of Economic Studies*, 87(2):997–1033.

Epstein, L. G. (2006). An axiomatic model of non-bayesian updating. *The Review of Economic Studies*, 73(2):413–436.

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

Feddersen, T. and Pesendorfer, W. (1997). Voting Behavior and Information Aggregation in Elections With Private Information. *Econometrica*, 65(5):1029–1058.

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.

Feddersen, T. J. and Pesendorfer, W. (1996). The swing voter's curse. *The American economic review*, pages 408–424.

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

Ginzburg, B., Guerra, J.-A., and Lekfuangfu, W. N. (2022). Counting on my vote not counting: Expressive voting in committees. *Journal of Public Economics*, 205:104555.

Gneezy, U. (2005). Deception: The role of consequences. *American Economic Review*, 95(1):384–394.

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.

Hadfield, J. D. (2010). MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. *Journal of Statistical Software*, 33(2):1–22.

Kahneman, D. and Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, 80(4):237–251.

Le Quement, M. T. and Marcin, I. (2020). Communication and voting in heterogeneous committees: An experimental study. *Journal of Economic Behavior & Organization*, 174:449–468.

Mandler, M. (2012). The fragility of information aggregation in large elections. *Games and Economic Behavior*, 74(1):257–268.

Massari, S. (2021). Price probabilities: A class of bayesian and non-bayesian prediction rules. *Economic Theory*, 72(1):133–166.

McKelvey, R. D. and Palfrey, T. R. (1998). Quantal response equilibria for extensive form games. *Experimental Economics*, 1(1):9–41.

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

Midjord, R., Rodríguez Barraquer, T., and Valasek, J. (2021). When voters like to be right: An analysis of the condorcet jury theorem with mixed motives. *Journal of Economic Theory*, 198:1–25.

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

Ortoleva, P. (2012). Modeling the change of paradigm: Non-bayesian reactions to unexpected news. *American Economic Review*, 102(6):2410–2436.

Palley, A. B. and Soll, J. B. (2019). Extracting the wisdom of crowds when information is shared. *Management Science*, 65(5):2291–2309.

Sánchez-Pagés, S. and Vorsatz, M. (2007). An experimental study of truth-telling in a sender–receiver game. *Games and Economic Behavior*, 61(1):86–112.

Sandroni, A., Epstein, L. G., and Noor, J. (2008). Non-bayesian updating: A theoretical framework. *Theoretical Economics*, 3(2):193–229.

Tversky, A. and Kahneman, D. (1982). Evidential impact of base rates. In Kahneman, D., Slovic, P., and Tversky, A., editors, *Judgment under Uncertainty: Heuristics and Biases*, pages 153–160. Cambridge University Press, Cambridge.

A Supplementary technical material

A.1 Companion paper and shared experimental data

This paper and Breitmoser and Valasek (2024) use the same experimental dataset. The companion paper studies how the choice of voting rule (Majority versus Unanimity) affects outcomes in the presence of expressive incentives. The present paper instead uses the joint distribution of private signals, messages, and votes to diagnose which departures from Bayesian equilibrium are empirically useful for organizing committee behavior. In particular, our identifying discipline comes from fitting communication and voting simultaneously, and from testing structural restrictions on beliefs and on the mapping from messages to outcomes.

A.2 Exact model definition, belief distortion, and equilibrium objects

Game form. The environment is the voting game Γ in Section 2.1. There are $N = 3$ players, a binary state $\omega \in \{R,B\}$, conditionally independent private signals $s_i \in \{R,B\}$ with $\Pr(s_i = \omega \mid \omega) = \alpha \in (1/2,1)$, and common prior $\Pr(\omega = R) = 1/2$. After observing s_i, each player sends a binary message $m_i \in \{R,B\}$ and then, after observing the message profile, submits a binary vote $v_i \in \{R,B\}$. The voting rule $D \in \{\text{Majority}, \text{Unanimity}\}$ maps votes into a committee decision $X \in \{R,B\}$; for Unanimity we use the reduced-form default-R convention described in Section 2.1. Payoffs are given by (2).

A (pure or mixed) strategy for player i is a pair (σ_i, τ_i), where $\sigma_i(\cdot)$ maps signals into distributions over messages and $\tau_i(\cdot)$ maps voting-stage information sets into distributions over votes. Throughout we restrict attention to symmetric strategies (σ, τ) and use the sufficient-statistic representation from the main text: $\sigma(s)$ is the probability of sending R after signal s, and $\tau(s, m_i, M)$ is the probability of voting R after signal s, own message m_i, and $M = \#\{j : m_j = B\}$.

Belief distortion. Let T_κ be the log-odds distortion defined in (3). Given a strategy profile and an information set I, write

$$p^B(I) \;\equiv\; \Pr(\omega = R \mid I)$$

for the Bayesian posterior induced by the objective model and the strategy profile. The distorted posterior is

$$p^\kappa(I) \;\equiv\; T_\kappa\big(p^B(I)\big). \tag{22}$$

We assume $\kappa > 0$ is common knowledge and that agents evaluate expected payoffs using $p^\kappa(\cdot)$ rather than $p^B(\cdot)$.

Sequential optimality with distorted beliefs. Fix $\kappa > 0$. At each information set, continuation payoffs are computed under the distorted posterior (22) and the continuation strategies and voting rule.

Definition 8 (Distorted sequential equilibrium)**.** *Fix $\kappa > 0$. A* distorted sequential equilibrium *is a symmetric strategy profile (σ, τ) together with a system of Bayesian posteriors $p^B(\cdot)$ such that: (i) $p^B(\cdot)$ is Bayes-consistent with (σ, τ) on the equilibrium path; (ii) at every information set, prescribed actions maximize expected utility when beliefs are evaluated using the distorted posterior system $p^\kappa(\cdot)$ defined by* (22)*.*

Definition 9 (Distorted logit QRE)**.** *Fix (κ, λ) with $\kappa > 0$ and $\lambda > 0$. A* distorted logit QRE *is a symmetric strategy profile (σ, τ) such that at every information set I, each feasible action a is chosen*

with probability proportional to $\exp\{\lambda EU^{\kappa}(a \mid I)\}$*, where* $EU^{\kappa}(a \mid I)$ *is the continuation expected utility computed under distorted beliefs* $p^{\kappa}(\cdot)$ *and continuation play given by* (σ,τ).

Equilibrium selection (when multiple equilibria exist). When the benchmark model admits multiple symmetric equilibria under the experimental parameters, we select among them as follows. First, we restrict attention to symmetric equilibria. Second, we select equilibria that maximize a representative agent's ex ante expected payoff (under the common prior and symmetry). Third, if ties remain, we select the limit point of the (appropriate) distorted logit QRE as $\lambda \to \infty$ in the sense of Definition 9.

A.3 Expected payoffs in messaging and voting information sets

This subsection records the probability objects used to compute continuation payoffs under a given symmetric strategy profile (σ,τ).

Induced distribution over terminal histories. Let $h = (\omega, s, m, v)$ denote a terminal history, where $s = (s_1, s_2, s_3)$, $m = (m_1, m_2, m_3)$, and $v = (v_1, v_2, v_3)$. Conditional independence of signals implies

$$\Pr(s \mid \omega) \;=\; \prod_{i=1}^{3} \Pr(s_i \mid \omega), \qquad \Pr(s_i = \omega \mid \omega) = \alpha,\ \Pr(s_i \neq \omega \mid \omega) = 1-\alpha.$$

Given (σ,τ), the induced probability of a terminal history is

$$\Pr_{\sigma,\tau}(h) \;=\; \Pr(\omega) \prod_{i=1}^{3} \Pr(s_i \mid \omega) \prod_{i=1}^{3} \Pr_{\sigma}(m_i \mid s_i) \prod_{i=1}^{3} \Pr_{\tau}\big(v_i \mid s_i, m_i, M(m)\big), \tag{23}$$

where $M(m) = \#\{j : m_j = B\}$ and $\Pr(\omega = R) = \Pr(\omega = B) = 1/2$. The realized committee decision is $X = X(v)$ as determined by the voting rule D.

Voting-stage continuation payoff. Fix player i and a voting-stage information set $I_i^V = (s_i, m_i, M)$, where M is the total number of B messages in the committee. For a contemplated vote $v_i \in \{R, B\}$, the (Bayesian) continuation payoff is

$$EU_i^B(v_i \mid I_i^V) \;=\; \sum_{\omega} \sum_{s_{-i}} \sum_{m_{-i}\,:\,M(m_i,m_{-i})=M} \sum_{v_{-i}} \pi_i\big(X(v_i, v_{-i}), \omega, v_i\big)\, \Pr_{\sigma,\tau}\big(\omega, s_{-i}, m_{-i}, v_{-i} \mid I_i^V, v_i\big), \tag{24}$$

where the conditional probability is induced by (23) given (s_i, m_i) and the restriction on M. Under κ-distortion, the same object is computed with $\Pr(\omega = \cdot \mid I_i^V)$ replaced by the distorted posterior $p^{\kappa}(I_i^V)$ in (22). The voting best-response condition and the pivotality formulation in Section 3.3 follow by comparing $EU_i^{\kappa}(B \mid I_i^V)$ and $EU_i^{\kappa}(R \mid I_i^V)$.

Messaging-stage continuation payoff. Fix player i and a messaging-stage information set given by the private signal s_i. For a contemplated message $m_i \in \{R, B\}$, the continuation payoff integrates over the subsequent message profile, votes, and outcomes:

$$EU_i^B(m_i \mid s_i) \;=\; \sum_{\omega} \sum_{s_{-i}} \sum_{m_{-i}} \sum_{v} \pi_i\big(X(v), \omega, v_i\big)\, \Pr_{\sigma,\tau}\big(\omega, s_{-i}, m_{-i}, v \mid s_i, m_i\big), \tag{25}$$

with the conditional probability again induced by (23). Under κ-distortion, the posterior $\Pr(\omega = \cdot \mid s_i)$ entering the conditional distribution is replaced by its distorted counterpart $p^{\kappa}(s_i)$, as in (22). The

payoff differences $dEP(s)$ used in Section 3.2 and in (20) are obtained from (25) by taking the difference between $m_i = B$ and $m_i = R$.

Computational remark. Because $N = 3$ and all variables are binary, the sums in (24)–(25) are finite and small. Under symmetry and conditional independence, many terms can be grouped by the sufficient statistics used in the main text (own signal, own message, and the total number of B messages), which is the basis for the numerical prediction and estimation routines reported in the Appendix.

A.4 Proofs of theoretical predictions

This appendix collects formal results underlying the theoretical "predictions" used in the main text. Throughout, the primitives and equilibrium objects are as in Sections 2.1–2.2 and Appendix A.2. In particular, $N = 3$, $\omega \in \{R, B\}$, signals satisfy $\Pr(s_i = \omega \mid \omega) = \alpha \in (1/2, 1)$ with common prior $\Pr(\omega = R) = 1/2$, and payoffs are given by (2).

Selection for point predictions. When we report a single "benchmark" prediction (for either Bayesian or distorted beliefs), we refer to the limiting-logit selection described in Appendix A.7. For convenience we state the object here.

Definition 10 (Limiting-logit point prediction)**.** *Fix a behavioral specification (Bayesian beliefs or distorted beliefs) and the extensive-form logit response defined in Appendix A.2 and Definition 9. For each precision $\lambda > 0$, let $(\sigma^\lambda, \tau^\lambda)$ denote a symmetric logit-QRE of the induced game. A* limiting-logit point prediction *is any accumulation point of $(\sigma^\lambda, \tau^\lambda)$ along a sequence $\lambda_n \to \infty$. When multiple accumulation points exist, we report the limit selected by the numerical homotopy procedure in Appendix A.7.*

Majority: voting incentives and a useful threshold

The following lemma records a convenient decomposition: under Majority, the only way a vote affects the committee outcome is through pivotality, whereas the expressive payoff is obtained whenever one votes R, independently of pivotality.

Lemma 2 (Voting incentives under Majority)**.** *Fix a voting-stage information set I for player i under Majority. Let*

$$p(I) \equiv \Pr(\omega = B \mid I) \qquad \text{and} \qquad \pi(I) \equiv \Pr(i \text{ is pivotal} \mid I),$$

where pivotal means that the committee outcome equals B if i votes B and equals R if i votes R. Holding continuation play fixed, the expected payoff difference between voting B and voting R satisfies

$$EU(B \mid I) - EU(R \mid I) = \pi(I)\, C\,(2p(I) - 1) - K. \tag{26}$$

In particular, if $\pi(I) = 1$ (pivotality is certain at I), then voting B is optimal if and only if $p(I) > \frac{1}{2}(1 + K/C)$.

Proof. If i is not pivotal at I, the committee outcome is the same under either vote, so the only payoff difference comes from the expressive term: voting R yields K and voting B yields 0, giving $EU(B) - EU(R) = -K$ in non-pivotal states. If i is pivotal, the committee outcome equals B under vote B and equals R under vote R, so the common-value payoff difference is $C \cdot \mathbf{1}\{\omega = B\} - C \cdot \mathbf{1}\{\omega = R\} = C(2\mathbf{1}\{\omega = B\} - 1)$, with expectation $C(2p(I) - 1)$; the expressive difference remains

$-K$. Taking expectations and weighting by $\pi(I)$ yields (26). The final claim follows by setting $\pi(I) = 1$. □

Lemma 3 (Pivotal-vote threshold under Majority)**.** *Fix a voting-stage information set at which player i is pivotal in the sense that the committee outcome equals B if she votes B and equals R if she votes R. Let $p \equiv \Pr(\omega = B)$ denote her belief at that information set. Under Majority, voting B is optimal if and only if*

$$p > p^\star \equiv \frac{1}{2}\Big(1 + \frac{K}{C}\Big). \tag{27}$$

For the experimental calibrations, $p^\star = 0.625$ in the Low treatment $(C,K) = (40,10)$ and $p^\star = 5/7 \approx 0.7143$ in the High treatment $(C,K) = (35,15)$.

Proof. This is the special case $\pi(I) = 1$ of Lemma 2. The numerical values follow by substitution. □

Benchmark Bayesian equilibrium under Majority

Proposition 2 (Benchmark equilibrium: High treatment under Majority)**.** *In the High treatment $(C,K) = (35,15)$ under Majority, there exists a symmetric sequential equilibrium in which all players vote R at every voting information set. In that equilibrium, messages are payoff-irrelevant on path; under the limiting-logit selection of Definition 10, the corresponding on-path message behavior is babbling, $\sigma(R) = \sigma(B) = 1/2$.*

Proof. Consider the voting strategy $\tau \equiv 1$, i.e. every player votes R with probability one at every voting information set. Given $\tau \equiv 1$, the committee outcome is $X = R$ regardless of messages. Hence, at any voting information set, deviating from R to B cannot affect X and strictly lowers payoff by forfeiting the expressive bonus K. Therefore $\tau \equiv 1$ is sequentially rational for any beliefs.

Given $\tau \equiv 1$, messages do not affect any player's payoff on path, so any message strategy is sequentially rational on path; beliefs can be completed off path to obtain a symmetric sequential equilibrium.

Under the limiting-logit selection, equal continuation payoffs from the two messages imply that the logit best response assigns probability $1/2$ to each message at every on-path messaging information set. Therefore $\sigma(R) = \sigma(B) = 1/2$ at the selected limit. □

Lemma 4 (Existence of limiting-logit predictions in finite games)**.** *Fix any behavioral specification (Bayesian or distorted) and any treatment. For each $\lambda > 0$, a symmetric logit-QRE exists. Moreover, any sequence of symmetric logit-QREs $(\sigma^{\lambda_n}, \tau^{\lambda_n})$ with $\lambda_n \to \infty$ has a convergent subsequence, and any accumulation point is a symmetric sequential equilibrium of the corresponding behavioral game.*

Proof. For each $\lambda > 0$, the logit best-response correspondence maps a product of finite simplices into itself and is continuous (indeed single-valued) given the finiteness of the game tree and the continuity of expected payoffs in strategies. Existence of a fixed point follows from Brouwer's theorem, and symmetry can be imposed by restricting attention to the symmetric subspace.

Compactness of the product of simplices implies any sequence has a convergent subsequence. Standard arguments for logit response in finite extensive-form games imply that as $\lambda \to \infty$, logit best responses place vanishing probability on strictly suboptimal actions at any reached information set, so any accumulation point is sequentially rational given the induced beliefs; with beliefs completed by Bayes' rule on path, this yields a symmetric sequential equilibrium. □

Proposition 3 (Benchmark equilibrium: Low treatment under Majority (selected numerical prediction))**.** *In the Low treatment* $(C,K) = (40,10)$ *under Majority, the benchmark equilibrium* point prediction *reported in the main text is the limiting-logit point prediction of Definition 10 under Bayesian beliefs. The resulting strategy profile is a symmetric sequential equilibrium (Lemma 4), and its numerical values are computed as described in Appendix A.7.*

Proof. For each $\lambda > 0$, the symmetric Bayesian logit-QRE exists by Lemma 4. The point prediction is, by definition, the accumulation point selected by the numerical procedure in Appendix A.7. That accumulation point is a symmetric sequential equilibrium by Lemma 4. The numerical values stated in the main text are those returned by the algorithm. □

Truthful-messaging benchmark (lying aversion) under Majority

Proposition 4 (Truthful messaging benchmark under Majority)**.** *Assume truthful messaging is exogenously imposed, i.e.* $m_i = s_i$ *for all i (the "lying aversion" benchmark). Under Majority:*

1. *In the High treatment* $(C,K) = (35,15)$*, voting R strictly dominates voting B at every voting information set; hence the unique symmetric equilibrium outcome has all agents voting R.*
2. *In the Low treatment* $(C,K) = (40,10)$*, voting R is optimal at every voting information set except when* $(s_1,s_2,s_3) = (B,B,B)$*. At that information set there is a unique symmetric mixed equilibrium in which each player votes B with probability*

$$q^\star \;=\; \frac{1}{2} + \frac{\sqrt{114}}{76} \approx 0.6405, \tag{28}$$

and votes R with probability $1 - q^\star \approx 0.3595$.

Proof. Under truthful messaging, the message profile reveals the full signal profile. Hence the voting stage following any realized (s_1,s_2,s_3) is a complete-information voting game in which all players share the same posterior $p = \Pr(\omega = B \mid s_1,s_2,s_3)$.

Let $p_{BBB} \equiv \Pr(\omega = B \mid B,B,B)$. Bayes' rule yields

$$p_{BBB} = \frac{\alpha^3}{\alpha^3 + (1-\alpha)^3}. \tag{29}$$

For $\alpha = 0.6$, $p_{BBB} = 27/35$.

Step 1 (High treatment). Fix any signal profile and any conjecture about the other two votes. Under Majority, $\pi(I) \le 1/2$ at any information set I in a symmetric 3-player voting subgame, because pivotality requires a 1–1 split among the other two votes. By Lemma 2,

$$EU(B \mid I) - EU(R \mid I) \;\le\; \frac{1}{2} C\,(2p_{BBB} - 1) - K.$$

In the High treatment, substituting $C = 35$, $K = 15$, and $p_{BBB} = 27/35$ yields $\frac{1}{2} \cdot 35 \cdot (19/35) - 15 = 9.5 - 15 < 0$. Hence $EU(B \mid I) < EU(R \mid I)$ at every voting information set, so voting R strictly dominates voting B. This proves 1.

Step 2 (Low treatment). In the Low treatment, p is maximized at BBB, so if voting B is not optimal at BBB it cannot be optimal elsewhere. Consider the complete-information voting game at BBB and let q denote the common probability with which each player votes B there. If player i votes B, the committee selects B unless both other players vote R (probability $(1-q)^2$). If player i votes

R, the committee selects B only if both other players vote B (probability q^2). Using payoffs (2), indifference $EU(B) = EU(R)$ reduces to

$$2Cq(1-q)(2p_{BBB}-1) = K.$$

Substituting $(C,K) = (40,10)$ and $p_{BBB} = 27/35$ yields $152q^2 - 152q + 35 = 0$, whose two roots are $q = \frac{1}{2} \pm \frac{\sqrt{114}}{76}$. The unique symmetric equilibrium with $q \geq 1/2$ is (28). At any other signal profile, the posterior is strictly smaller than p_{BBB}, so the same inequality implies $EU(B \mid I) < EU(R \mid I)$ for all $q \in [0,1]$, and thus voting R is optimal. This proves 2. □

Belief distortion and monotone comparative statics

For convenience we restate the distortion mapping in the notation used in the main text.

Definition 11 (Log-odds distortion)**.** *For $\kappa > 0$, define $T_\kappa : [0,1] \to [0,1]$ by*

$$T_\kappa(p) \;=\; \frac{p^\kappa}{p^\kappa + (1-p)^\kappa}. \tag{30}$$

Lemma 5 (Log-odds scaling and extremeness)**.** *For any $p \in (0,1)$,*

$$\log\Big(\frac{T_\kappa(p)}{1-T_\kappa(p)}\Big) = \kappa \log\Big(\frac{p}{1-p}\Big).$$

Moreover, if $p > 1/2$ then $T_\kappa(p)$ is strictly increasing in κ and $\lim_{\kappa\to\infty} T_\kappa(p) = 1$; if $p < 1/2$ then $T_\kappa(p)$ is strictly decreasing in κ and $\lim_{\kappa\to\infty} T_\kappa(p) = 0$; and $\lim_{\kappa\to 0} T_\kappa(p) = 1/2$ for all $p \in (0,1)$.

Proof. By (30),

$$\frac{T_\kappa(p)}{1-T_\kappa(p)} = \frac{p^\kappa}{(1-p)^\kappa} = \Big(\frac{p}{1-p}\Big)^\kappa,$$

and taking logs yields the scaling identity. Monotonicity in κ follows because $\frac{p}{1-p} > 1$ if and only if $p > 1/2$, and limits follow from the fact that the log-odds diverge to $\pm\infty$ as $\kappa \to \infty$ and converge to 0 as $\kappa \to 0$. □

Lemma 6 (Monotone action shifts under log-odds distortion)**.** *Fix an information set I and two actions $a^\uparrow, a^\downarrow$. Let p denote the belief that the payoff-relevant state equals the state that favors $a^\uparrow$, i.e. $p = \Pr(\omega = \omega^\uparrow \mid I)$. Suppose that, holding continuation play fixed, the expected payoff difference*

$$\Delta(p) \;\equiv\; EU(a^\uparrow \mid p, I) - EU(a^\downarrow \mid p, I)$$

is (weakly) increasing in p. Then $\Delta(T_\kappa(p))$ is (weakly) increasing in κ whenever $p > 1/2$, and (weakly) decreasing in κ whenever $p < 1/2$. Under logit response with precision $\lambda > 0$, the choice probability of $a^\uparrow$ inherits the same monotonicity in κ.

Proof. By Lemma 5, $T_\kappa(p)$ moves monotonically in κ away from $1/2$ in the direction of p. Composing with the (weakly) increasing function $\Delta(\cdot)$ yields the first claim. Under logit response with precision $\lambda > 0$, $\Pr(a^\uparrow \mid I) = \big(1 + \exp\{-\lambda\Delta(\cdot)\}\big)^{-1}$ is strictly increasing in Δ, so the same monotonicity carries over. □

Overreaction predictions under Majority (numerical values)

The overreaction predictions reported in the main text for Majority are computed by solving the symmetric distorted logit-QRE fixed point for (σ, τ) (Definition 9) and then taking the limit as $\lambda \to \infty$ in the sense of Definition 10. The numerical procedure is described in Appendix A.7.

Unanimity: truth-telling in informative continuations

The Unanimity implementation in Section 2.1 selects B if and only if all votes are B; otherwise the enforced final vote (and the committee outcome) is R. This creates a natural veto property: any player can ensure $X = R$ by voting R.

Lemma 7 (Sufficient condition for B-type truth-telling under Unanimity)**.** *Fix Unanimity with default-R enforcement. Consider a continuation in which, at the voting stage, players vote B at the all-B message profile and vote R at all other message profiles. Suppose* $\alpha > 1/2$ *and*

$$p_{BBB} \equiv \Pr(\omega = B \mid B,B,B) > \frac{1}{2}\Big(1 + \frac{K}{C}\Big). \tag{31}$$

Fix a player i with signal $s_i = B$ *and suppose that, conditional on i sending* $m_i = B$*, the all-B message profile occurs with positive probability.*[9] *Then sending message* $m_i = B$ *is a strict best response at the messaging stage. For* $\alpha = 0.6$ *and* $(C,K) \in \{(40,10),(35,15)\}$*, condition* (31) *holds.*

Proof. Under the stated continuation, the committee selects B if and only if all three messages are B (and hence all votes are B). Fix a player with $s_i = B$.

If she sends $m_i = R$, the all-B message profile cannot occur and the outcome is $X = R$ surely, yielding expected payoff

$$C \Pr(\omega = R \mid s_i = B) + K.$$

If she sends $m_i = B$, then on the event that the other two messages are also B, the outcome becomes $X = B$ and her payoff equals $C \Pr(\omega = B \mid B,B,B) = C p_{BBB}$, whereas on the complement the outcome remains $X = R$ and her payoff equals $C \Pr(\omega = R \mid s_i = B, \cdot) + K$. Hence the expected payoff gain from sending B rather than R equals the probability of reaching the all-B message profile times

$$C p_{BBB} - \big(C(1 - p_{BBB}) + K\big) = C(2p_{BBB} - 1) - K,$$

which is strictly positive by (31) (equivalently Lemma 3). Under the stated positive-probability condition, this yields a strict gain, hence $m_i = B$ is a strict best response. The final claim follows because for $\alpha = 0.6$, $p_{BBB} = 27/35 > 0.7143 \geq \frac{1}{2}(1 + K/C)$ for both payoff calibrations. □

Lemma 8 (R-type veto messaging under Unanimity)**.** *Fix Unanimity with default-R enforcement and consider the same continuation as in Lemma 7. For any treatment and any* $\alpha \in (1/2, 1)$*, a player with signal* $s_i = R$ *weakly prefers sending* $m_i = R$ *to sending* $m_i = B$*, and strictly prefers* $m_i = R$ *whenever the all-B message profile would occur with positive probability conditional on* $m_i = B$*.*

Proof. If i sends $m_i = R$, the all-B message profile cannot occur and the outcome is $X = R$ surely, yielding $C \Pr(\omega = R \mid s_i = R) + K$. If instead she sends $m_i = B$, then whenever the all-B message profile occurs the outcome becomes $X = B$, in which case she forgoes the expressive bonus and obtains the common-value payoff only when $\omega = B$. Since $\Pr(\omega = B \mid s_i = R) < 1/2$ for $\alpha > 1/2$, the

[9] If the all-B message profile has zero probability conditional on $m_i = B$, then i is indifferent between messages under the stated continuation, so strictness cannot be concluded.

expression $C(2\Pr(\omega = B \mid s_i = R) - 1) - K$ is strictly negative, so reaching the all-B profile strictly lowers expected payoff; otherwise payoffs coincide. The conclusion follows. □

A.5 Portability: heterogeneous preference intensities and perceived signal precision

This appendix records a portability observation. Even when committee members have heterogeneous *intensities* for correctness across states, sufficiently high perceived signal precision can align induced preferences over the committee decision at all empirically relevant information sets. This restores an informative (truthful) equilibrium in a canonical cheap-talk committee game. The point is conceptual; we do not estimate this extension.

Environment. There are $N = 3$ committee members and a binary state $\omega \in \{R, B\}$ drawn from the common prior $\Pr(\omega = R) = 1/2$. Conditional on ω, each member i observes a private signal $s_i \in \{R, B\}$, independently across i, with objective accuracy $\Pr(s_i = \omega \mid \omega) = \alpha \in (1/2, 1)$. Members simultaneously send costless messages $m_i \in \{R, B\}$ and then vote $v_i \in \{R, B\}$. Under Majority rule, the committee decision is $X = R$ if at least two votes equal R and $X = B$ otherwise.

Members have (possibly heterogeneous) utilities $u_i(X, \omega)$ that satisfy correctness monotonicity:

$$u_i(\omega, \omega) > u_i(\bar{\omega}, \omega) \quad \text{for all } i \text{ and } \omega \in \{R, B\}, \tag{32}$$

where $\bar{R} = B$ and $\bar{B} = R$. Define state-specific gains from being correct,

$$\Delta_i^R \equiv u_i(R, R) - u_i(B, R) > 0, \qquad \Delta_i^B \equiv u_i(B, B) - u_i(R, B) > 0.$$

Perceived signal precision. Fix a perceived signal accuracy $\tilde{\alpha} \in (1/2, 1)$ that is common knowledge and is used for belief formation. In the main text, $\tilde{\alpha}$ can be generated by the log-odds scaling distortion with parameter κ; equivalently,

$$\tilde{\alpha}(\kappa) = \frac{\alpha^\kappa}{\alpha^\kappa + (1 - \alpha)^\kappa}, \qquad \kappa > 0, \tag{33}$$

so that larger κ corresponds to higher "as-if" precision and $\tilde{\alpha}(\kappa) \uparrow 1$ as $\kappa \to \infty$.

Lemma 9. *Fix agent i and an information set I with posterior belief $\pi \equiv \Pr(\omega = R \mid I)$. Choosing $X = R$ maximizes i's expected utility if and only if*

$$\pi > \pi_i^\star \equiv \frac{\Delta_i^B}{\Delta_i^R + \Delta_i^B}. \tag{34}$$

Proof. Compute the expected-utility difference:

$$EU_i(R \mid I) - EU_i(B \mid I) = \pi \Delta_i^R - (1 - \pi)\Delta_i^B.$$

This is positive if and only if $\pi > \Delta_i^B / (\Delta_i^R + \Delta_i^B)$. □

A sufficient condition for an informative equilibrium. Under truthful messages, the message profile reveals the signal profile. With $N = 3$ and prior $1/2$, the perceived posterior after observing a *majority* of R signals (two R and one B) equals $\tilde{\alpha}$, and after observing a majority of B signals equals

$1-\tilde{\alpha}$.[10]

Let

$$\bar{\pi}^\star \equiv \max_i \pi_i^\star, \qquad \underline{\pi}^\star \equiv \min_i \pi_i^\star.$$

Proposition 5. *Suppose* $\tilde{\alpha}$ *satisfies*

$$\tilde{\alpha} > \bar{\pi}^\star \quad \textit{and} \quad 1-\tilde{\alpha} < \underline{\pi}^\star. \tag{35}$$

Then the following strategy profile is a perfect Bayesian equilibrium under Majority rule: (i) each agent sends the truthful message $m_i = s_i$; (ii) after observing messages, each agent votes for the alternative supported by the majority of messages (equivalently, for the majority of revealed signals). In this equilibrium, the committee implements the majority of signals.

Proof. Fix beliefs induced by truth-telling. If the revealed signal majority is R, then the perceived posterior is $\pi = \tilde{\alpha} > \bar{\pi}^\star \geq \pi_i^\star$ for all i, so Lemma 9 implies every agent strictly prefers $X = R$. If the revealed signal majority is B, then $\pi = 1-\tilde{\alpha} < \underline{\pi}^\star \leq \pi_i^\star$ for all i, so every agent strictly prefers $X = B$. Hence voting with the message majority is a best response for every agent at every information set reached under truthful messages.

At the messaging stage, under Majority rule and the prescribed voting behavior, agent i's message affects the committee decision only when the other two messages are split (one R and one B). In that event the committee decision equals i's message. If $s_i = R$, the perceived posterior that $\omega = R$ equals $\tilde{\alpha} > \bar{\pi}^\star \geq \pi_i^\star$, so by Lemma 9 agent i strictly prefers $X = R$ and hence strictly prefers sending $m_i = R$ when pivotal. If $s_i = B$, the perceived posterior equals $1-\tilde{\alpha} < \underline{\pi}^\star \leq \pi_i^\star$, so agent i strictly prefers $X = B$ and hence strictly prefers sending $m_i = B$ when pivotal. Off the pivotal event, the message does not affect X, so truth-telling is weakly optimal. This establishes sequential rationality at the messaging stage and completes the equilibrium construction. □

Interpretation. Proposition 5 provides a tractable sufficient condition under which increased responsiveness to private information (captured by high perceived precision $\tilde{\alpha}$, or equivalently large κ in (33)) aligns induced preferences over the committee decision even when agents differ in the relative intensity with which they value correctness across states. The result does not claim that belief distortions are the only route to informative equilibria under heterogeneous preferences; it simply illustrates how "as-if" high precision reduces scope for strategic misreporting driven by preference conflicts under mixed evidence. This appendix also highlights why the expressive-payoff environment in the main text is nontrivial: with expressive incentives, truthful revelation need not eliminate the voting-stage collective-action problem, whereas in the present reduced-form environment (without expressive payoffs) sufficiently extreme perceived precision can align incentives over the committee decision itself.

A.6 Messages-only diagnostic: state-contingent payoffs and implied state beliefs

This appendix records an auxiliary diagnostic used for interpretation in the main text. Conditional on the true state, the realized continuation payoff from sending the message that matches the state is substantially higher than the payoff from sending the opposite message (Table 8). This payoff structure implies that any belief formation that overweights $\omega = s$ after observing a signal s makes truthful messaging *instrumentally* attractive. At the same time, message data alone do not distinguish

[10] For example, with two R and one B, the likelihood ratio in favor of R equals $\tilde{\alpha}/(1-\tilde{\alpha})$, hence the posterior equals $\tilde{\alpha}$ under prior $1/2$.

belief distortions from nonstandard preferences over honesty, which is why the main text relies on voting behavior for identification.

State-contingent continuation payoffs from messages. Let $\bar{\pi}(m,\omega)$ denote the average realized monetary payoff to a sender who chooses message $m \in \{R,B\}$ when the realized state is $\omega \in \{R,B\}$, integrating over subsequent play observed in the data. Define the state-contingent payoff difference

$$\Delta\pi(\omega) \equiv \bar{\pi}(B,\omega) - \bar{\pi}(R,\omega). \tag{36}$$

Table 8 reports $\bar{\pi}(m,\omega)$ by treatment. In all treatments, $\bar{\pi}(R,R) > \bar{\pi}(B,R)$ and $\bar{\pi}(B,B) > \bar{\pi}(R,B)$, i.e. the message matching the state yields a large payoff advantage under realized continuation play.

Expected payoff differences given a belief about the state. Fix a sender who has received signal $s \in \{R,B\}$ and holds belief $\pi_s \equiv \Pr(\omega = R \mid s)$. The expected payoff advantage of message B over message R is

$$\Delta EP(s) \equiv EP(B \mid s) - EP(R \mid s) = \pi_s \Delta\pi(R) + (1-\pi_s)\Delta\pi(B). \tag{37}$$

Since $\Delta\pi(R) < 0$ and $\Delta\pi(B) > 0$ in Table 8, increasing π_s makes B-messaging less attractive after $s = R$ and more attractive after $s = B$; equivalently, overweighting $\omega = s$ tends to increase the perceived profitability of the truthful message.

Why messages alone are not identifying. One may combine (37) with a logit choice rule to back out *implied* posteriors from observed message frequencies. We do not report that inversion here because (i) the paper's identification strategy does not rely on messages alone and (ii) honest-messaging preferences and belief distortions can generate similar message frequencies. The main text therefore disciplines beliefs using voting behavior conditional on message profiles, which imposes additional restrictions that a messages-only diagnostic cannot.

A.7 Computation of benchmark equilibria and limiting-logit predictions

This subsection documents how we compute the benchmark equilibrium predictions and the limiting-logit selections reported in the main text and tables. The goal is that a reader can reproduce the objects from the description below.

Strategy parametrization under symmetry. We restrict attention to symmetric behavioral strategies (σ,τ) as in the main text. The messaging strategy is summarized by the two probabilities

$$\sigma(R) \equiv \Pr(m_i = R \mid s_i = R), \qquad \sigma(B) \equiv \Pr(m_i = R \mid s_i = B).$$

The voting strategy is summarized by

$$\tau(s,m,M) \equiv \Pr(v_i = R \mid s_i = s,\ m_i = m,\ M = \#\{j : m_j = B\}),$$

where $M \in \{0,1,2,3\}$ counts the total number of B messages in the committee (including player i). This is the sufficient-statistic representation used throughout.

Expected utilities. Given (σ,τ), the induced probability of terminal histories is given in (23). Voting- and messaging-stage continuation expected utilities are computed by finite enumeration of

terminal histories as in (24)–(25). In distorted-belief variants, the only modification is that the posterior over the state used in these conditional expectations is replaced by $p^{\kappa}(\cdot)$ as defined in (22); the objective signal process and continuation strategies remain unchanged.

Logit best-response operators. Fix (κ, λ) with $\kappa > 0$ and $\lambda > 0$. For each information set I and feasible action $a \in \{R, B\}$, define

$$BR^{\kappa,\lambda}(a \mid I) \;\equiv\; \frac{\exp\{\lambda EU^{\kappa}(a \mid I)\}}{\exp\{\lambda EU^{\kappa}(R \mid I)\} + \exp\{\lambda EU^{\kappa}(B \mid I)\}}.$$

A symmetric distorted logit QRE (Definition 9) is a fixed point (σ, τ) of these operators across all information sets.

Numerical fixed-point computation. For each parameter configuration (rule D, payoffs (C, K), signal accuracy α, distortion κ, and precision λ), we compute a symmetric fixed point as follows.

Step 1: initialization. Start from an interior strategy $(\sigma^{(0)}, \tau^{(0)})$ with all components in $(0, 1)$. To address potential multiplicity, we use multiple initializations (including near-babbling and near-truthful message profiles and several voting seeds).

Step 2: policy evaluation. Given $(\sigma^{(t)}, \tau^{(t)})$, compute $EU^{\kappa}(\cdot \mid I)$ for all information sets by enumerating terminal histories using (23) and the conditional expectations (24)–(25).

Step 3: logit update with damping. Update each component by a damped best response:

$$(\sigma^{(t+1)}, \tau^{(t+1)}) \;=\; (1-\eta)\,(\sigma^{(t)}, \tau^{(t)}) \;+\; \eta\, BR^{\kappa,\lambda}(\cdot \mid \cdot),$$

with step size $\eta \in (0, 1]$ chosen to ensure contraction (we use a small η when oscillations are detected).

Step 4: convergence and verification. Iterate until the sup-norm distance between successive iterates falls below a tolerance. We additionally verify that the resulting strategy solves the fixed-point equations up to tolerance by checking the maximum absolute residual across all information sets.

Limiting-logit selection. When the main text reports a "limiting-logit" prediction, we approximate the $\lambda \to \infty$ selection by evaluating the fixed point on an increasing grid of large λ values and checking stability of (i) the qualitative pattern of play and (ii) the numerical values of the key strategy components. The reported strategy is the largest-λ fixed point for which stability is attained. This procedure implements the selection device described in Appendix A.2.

Efficiency and welfare calculations. Given any computed strategy profile (σ, τ), we compute (i) ex ante expected payoff (under the common prior and symmetry) and (ii) the "efficiency" statistic reported in the paper. The latter is the probability that the committee decision coincides with the majority of private signals:

$$c(\sigma, \tau) \;\equiv\; \Pr\big(X = \mathrm{maj}(s_1, s_2, s_3)\big),$$

where $\mathrm{maj}(\cdot)$ is the majority operator on $\{R, B\}^3$. Both objects are computed by summing (23) over terminal histories that satisfy the relevant event.

A.8 Detailed predictions: efficient limiting-logit equilibria

This subsection clarifies how the point predictions reported in the main text are selected when multiple symmetric equilibria exist.

Benchmark (Bayesian) case. For the benchmark model we set $\kappa = 1$. If the symmetric limiting-logit selection yields a unique prediction, we report its strategy components (σ, τ) and the implied outcome statistics (e.g. $c(\sigma, \tau)$). If multiple symmetric limiting-logit candidates arise from different initializations, we select the equilibrium that maximizes ex ante expected payoff under the objective model and symmetry. When ties remain, we select the equilibrium that is the stable limit of the logit fixed points as λ increases along the grid used in Appendix A.7.

Belief-distorted (overreaction) case. For the belief-distorted model we fix $\kappa \neq 1$ and compute distorted logit QRE fixed points as above. The selection criterion is the same: among symmetric candidates, we report the equilibrium that maximizes ex ante expected payoff, with ties broken by the stable limiting-logit path as λ increases.

Perfect-overreaction benchmark. When we report the "perfect overreaction" benchmark corresponding to $\kappa \to \infty$, we approximate it by evaluating the distorted model at a sufficiently large κ and verifying that the implied strategy components and outcome statistics are numerically stable to further increases in κ. The reported objects should therefore be read as an accurate numerical approximation to the $\kappa \to \infty$ limit.

B Experimental Instructions (original)

Instruktionen

Dies ist ein Experiment zur Entscheidungsfindung. Vielen Dank für Ihre Teilnahme!

Bitte lesen Sie diese Instruktionen sorgfältig. Es ist wichtig, dass Sie während des gesamten Experimentes nicht mit anderen Teilnehmer kommunizieren. Falls Sie Fragen habe, lesen Sie bitte noch einmal in diesen Instruktionen nach. Bei weiteren Fragen melden Sie sich bitte. Wir werden dann zu Ihnen kommen und die Fragen persönlich beantworten. Bitte fragen Sie nicht laut.

Das gesamte Experiment läuft über die Computer Terminals, und jedwede Interaktion zwischen Ihnen wird über die Computer laufen. Sie werden für Ihre Teilnahme am Ende des Experimentes in bar bezahlt. Unterschiedliche Teilnehmer werden unterschiedliche Beträge verdienen. Ihr Verdienst hängt sowohl von Ihren Entscheidungen ab als auch von den Entscheidungen anderer Teilnehmer und Zufall.

Das Experiment läuft über 50 Runden. Die Regeln sind über alle Runden und für alle Teilnehmer dieselben. Zu Beginn jeder Runde werden Sie zufällig in Gruppen aus drei Teilnehmern eingeteilt. In jeder Runde werden Sie nur mit den Teilnehmern in Ihrer Gruppe interagieren. Sie werden die Identität der anderen Teilnehmer in Ihrer Gruppe nicht erfahren. Wir werden die anderen Teilnehmern in Ihrer Gruppe als "Teilnehmer 2" und "Teilnehmer 3" bezeichnen, aber beachten Sie, dass nach jeder Runde die Gruppen neu eingeteilt werden. Ihre Gruppe wird eine Entscheidung basierend auf einer Abstimmung fällen. Diese Entscheidung ist einfach die Wahl zwischen zwei Urnen, der blauen Urne und der roten Urne. Das genaue Prozedere erklären wir Ihnen im Folgenden.

Table 13: The sequential equilibria that can be represented as limiting logit equilibria in the four treatment

(a) Majority: $N = 3, \quad P = 1, \quad K = 10/40, \quad \alpha = 0.6$

	$\sigma(R)$	$\sigma(B)$	$\tau(R,B,0-2)$			$\tau(B,R,0-2)$			$\tau(B,B,0-2)$			π	c
Equilibrium	1	0.56	1	1	1	1	1	1	1	0	1	0.7696	0.6165
Lying Aversion	1	0	1	1	1	1	1	1	1	1	0.36	0.7811	0.5987
Overreaction	1	0.1	1	1	1	1	1	1	1	0	0.15	1.102	0.9532
Level-K (1)	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.625	0.5
	0.5	0.5	1	1	1	1	1	1	1	1	1	0.75	0.5
	0.5	0.5	1	1	1	1	1	1	1	1	1	0.75	0.5
	0.5	0.5	1	1	1	1	1	1	1	1	1	0.75	0.5
Level-L (1)	1	0	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.625	0.5
	1	0	1	1	1	1	1	0	1	1	0	0.791	0.64
	0	0	1	1	1	1	1	1	1	1	1	0.7495	0.5
	1	0	1	1	1	1	1	1	1	1	1	0.75	0.5

(b) Majority: $N = 3, \quad P = 1, \quad K = 15/35, \quad \alpha = 0.6$

	$\sigma(R)$	$\sigma(B)$	$\tau(R,B,0-2)$			$\tau(B,R,0-2)$			$\tau(B,B,0-2)$			π	c
Equilibrium	0.5	0.5	1	1	1	1	1	1	1	1	1	0.9286	0.5
Lying Aversion	1	0	1	1	1	1	1	1	1	1	1	0.9286	0.5
Overreaction	1	0.28	1	1	1	1	1	1	1	0	0.31	1.1571	0.8446
Level-K (1)	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.7143	0.5
	0.5	0.5	1	1	1	1	1	1	1	1	1	0.9286	0.5
	0.5	0.5	1	1	1	1	1	1	1	1	1	0.9286	0.5
	0.5	0.5	1	1	1	1	1	1	1	1	1	0.9286	0.5
Level-L (1)	1	0	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.7143	0.5
	1	0	1	1	1	1	1	1	1	1	1	0.9286	0.5
	1	0	1	1	1	1	1	1	1	1	1	0.9286	0.5
	1	0	1	1	1	1	1	1	1	1	1	0.9286	0.5

(c) Unanimity: $N = 3, \quad P = 1, \quad K = 10/40, \quad \alpha = 0.6$

	$\sigma(R)$	$\sigma(B)$	$\tau(R,B,0-2)$			$\tau(B,R,0-2)$			$\tau(B,B,0-2)$			π	c
Equilibrium	0.5	0.5	1	1	1	0	0	0	0	0	0	0.791	0.64
Lying Aversion	1	0	1	1	1	0.99	1	0	0.95	0	0	0.791	0.64
Overreaction	1	0	1	1	0	1	0	0	1	0	0	1.114	1
Level-K (1)	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.6777	0.5
	0.5	0.5	1	1	1	1	1	1	1	1	1	0.75	0.5
	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.6777	0.5
	0.5	0.5	1	1	1	1	1	1	1	1	1	0.75	0.5
Level-L (1)	1	0	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.6777	0.5
	1	0	1	1	1	1	1	0	1	1	0	0.791	0.64
	1	0	0.5	0.5	1	0.5	0.5	0.5	0.5	0.5	0	0.7069	0.5995
	0	0	1	1	1	1	1	0	1	1	0	0.7905	0.64

(d) Unanimity: $N = 3, \quad P = 1, \quad K = 15/35, \quad \alpha = 0.6$

	$\sigma(R)$	$\sigma(B)$	$\tau(R,B,0-2)$			$\tau(B,R,0-2)$			$\tau(B,B,0-2)$			π	c
Equilibrium	0.5	0.5	1	1	1	0	0	0	0	0	0	0.9446	0.64
Lying Aversion	1	0	1	1	1	1	1	0	0.99	0	0	0.9446	0.64
Overreaction	1	0	1	1	0	1	0	0	0.97	0	0	1.2032	0.9999
Level-K (1)	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.8047	0.5
	0.5	0.5	1	1	1	1	1	1	1	1	1	0.9286	0.5
	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.8047	0.5
	0.5	0.5	1	1	1	1	1	1	1	1	1	0.9286	0.5
Level-L (1)	1	0	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.5	0.8047	0.5
	1	0	1	1	1	1	1	0	1	1	0	0.9446	0.64
	1	0	0.5	0.5	1	0.5	0.5	0.5	0.5	0.5	0	0.8161	0.5995
	0	0	1	1	1	1	1	0	1	1	0	0.9441	0.64

Die Urne. Es gibt zwei Urnen: die blaue Urne und die rote Urne. Die blaue Urne enthält 3 blaue Kugeln und 2 rote Kugeln. Die rote Urne enthält 3 rote Kugeln und 2 blaue Kugeln. Zu Beginn jeder Runde wird

Table 14: Theoretical predictions across treatments

	Messages		Voting											
	$\sigma(R)$	$\sigma(B)$	$\tau(R,R,0)$	$\tau(B,R,0)$	$\tau(R,R,1)$	$\tau(R,B,1)$	$\tau(B,B,1)$	$\tau(B,R,1)$	$\tau(R,R,2)$	$\tau(R,B,2)$	$\tau(B,B,2)$	$\tau(B,R,2)$	$\tau(R,B,3)$	$\tau(B,B,3)$
Majority 40-10														
Equilibrium	1	0.56	1	1	1	1	1	1	1	1	0	1	1	1
Lying Aversion	1	0	1	1	1	1	1	1	1	1	1	1	1	0.36
Overreaction	1	0.1	1	1	1	1	1	1	1	1	0	1	1	0.15
Level-K (1)	0.5	0.5	1	1	1	1	1	1	1	1	1	1	1	1
Level-L (1)	1	0	1	1	1	1	1	1	1	1	1	0	1	0
Majority 35-15														
Equilibrium	0.5	0.5	1	1	1	1	1	1	1	1	1	1	1	1
Lying Aversion	1	0	1	1	1	1	1	1	1	1	1	1	1	1
Overreaction	1	0.28	1	1	1	1	1	1	1	1	0	1	1	0.31
Level-K (1)	0.5	0.5	1	1	1	1	1	1	1	1	1	1	1	1
Level-L (1)	1	0	1	1	1	1	1	1	1	1	1	1	1	1
Unanimity 40-10														
Equilibrium	0.5	0.5	1	0	1	1	0	0	1	1	0	0	1	0
Lying Aversion	1	0	1	1	1	1	1	1	1	1	0	0	1	0
Overreaction	1	0	1	1	1	1	1	0	0	1	0	0	0	0
Level-K (1)	0.5	0.5	1	1	1	1	1	1	1	1	1	1	1	1
Level-L (1)	1	0	1	1	1	1	1	1	1	1	1	0	1	0
Unanimity 35-15														
Equilibrium	0.5	0.5	1	0	1	1	0	0	1	1	0	0	1	0
Lying Aversion	1	0	1	1	1	1	1	1	1	1	0	0	1	0
Overreaction	1	0	1	1	1	1	1	0	0	1	0	0	0	0
Level-K (1)	0.5	0.5	1	1	1	1	1	1	1	1	1	1	1	1
Level-L (1)	1	0	1	1	1	1	1	1	1	1	1	0	1	0

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

eine der Urnen zufällig gewählt. Diese Urne bezeichnen wir als gewählte Urne. Jede Urne wird mit gleicher Wahrscheinlichkeit gewählt, also jeweils mit 50% Wahrscheinlichkeit. Sie werden nicht erfahren welche Urne gewählt wurde bevor Sie Ihre Entscheidung treffen.

Die Kugel. Nachdem die Urne gewählt wurde, zeigt der Computer jedem von Ihnen eine Kugel, die zufällig aus der Urne Ihrer Gruppe gezogen wurde. Für jeden von Ihnen wird eine eigene Kugel gezogen ("mit Zurücklegen"). Jede Kugel in der Urne hat die gleiche Wahrscheinlichkeit, gezogen zu werden. Falls die blaue Urne für Ihre Gruppe gewählt wurde, ist für jeden von Ihnen die Wahrscheinlichkeit, eine blaue Kugel zu sehen, genau 60%, und die Wahrscheinlichkeit eine rote Kugel zu sehen ist 40%. Falls die rote Kugel für Ihre Gruppe gewählt wurde, ist für jeden von Ihnen die Wahrscheinlichkeit, eine blaue Kugel zu sehen, genau 40%, und die Wahrscheinlichkeit eine rote Kugel zu sehen ist 60%.

Die Nachricht. Nachdem jedem von Ihnen eine Kugel präsentiert wurde, kann jeder eine Nachricht senden. Die Nachricht ist entweder "rote Kugel" oder "blaue Kugel". Die Nachricht, die Sie senden, kann der Ihnen präsentierten Kugel gleichen, kann aber auch anders sein. Dies hängt von Ihrer Strategie und Ihren Präferenzen ab. Wenn alle Teilnehmer Ihre Nachricht versendet haben, werden Ihnen die Nachricht Ihrer Gruppe gezeigt. Da sie zu dritt sind, sieht jeder von Ihnen drei Nachrichten (inklusive der eigenen Nachricht), und jede dieser Nachrichten ist eine rote Kugel oder eine blaue Kugel.

Die Abstimmung [Mehrheit]. Nachdem Sie alle Nachrichten gesehen haben, erfolgt die Abstimmung. Sie können entweder für "rote Urne" oder "blaue Urne" stimmen. Ihre Stimme kann der Kugel, die Ihnen gezeigt wurde, oder der Nachricht, die sieht versendet haben, gleichen, muss aber nicht. Nur die Abstimmung zählt für Ihre Auszahlung.

Die Gruppenentscheidung ergibt sich aus der Mehrheitsregel. Falls mindestens zwei Teilnehmer Ihrer Gruppe (einschließlich Ihnen selbst) für die "rote Urne" stimmten, ist die Gruppenentscheidung "rote Urne". Falls mindestens zwei Teilnehmer für die "blaue Urne" stimmten, ist die Gruppenentscheidung "blaue Urne".

Auszahlung. Ihre Auszahlung pro Runde ergibt sich als Summe zweier Komponenten. Einerseits, wenn die Gruppenentscheidung mit der vom Computer gewählten Urne übereinstimmt, erhält jedes Mitglied Ihrer Gruppe 40 Taler. Wenn die Gruppenentscheidung nicht richtig ist, erhält jeder von Ihnen 0 Taler aus der Gruppenentscheidung. Andererseits, wenn Sie individuell für die "rote Urne" gestimmt haben, erhalten Sie persönlich zusätzlich 10 Taler. Wenn Sie für die "blaue Urne" stimmten, ist Ihr zusätzlicher Verdienst 0 Taler. Die folgenden Tabellen fassen dies noch einmal zusammen.

Der Computer wählte die **blaue Urne**

	Die Stimmen der anderen Gruppenmitglieder sind			
Ihre Stimme	Blau + Blau	Blau + Rot	Rot + Blau	Rot + Rot
Blaue Urne	40	40	40	0
Rote Urne	50	10	10	10

Der Computer wählte die **rote Urne**

	Die Stimmen der anderen Gruppenmitglieder sind			
Ihre Stimme	Blau + Blau	Blau + Rot	Rot + Blau	Rot + Rot
Blaue Urne	0	0	0	40
Rote Urne	10	50	50	50

Informationen am Ende jeder Runde. Sobald Sie und die anderen Teilnehmer abgestimmt haben, ist die Runde beendet. Zum Ende jeder Runde erhalten Sie die folgenden Informationen: Nachrichten und Stimmen aller Teilnehmer Ihrer Gruppe, die Gruppenentscheidung, die vom Computer gewählte Urne, Ihre Auszahlung.

Abschließender Verdienst. Am Ende des Experimentes werden die erworbenen Taler aller 50 Runden addiert und in Euro umgewandelt. Jeder Taler ist dann einen Cent wert. 100 Taler sind also 1 Euro wert. Zusätzlich erhalten Sie eine Basiszahlung von 5 Euro. Die Auszahlung erfolgt privat und für Sie ergibt sich keine Verpflichtung, anderen Ihren Verdienst mitzuteilen.

Die Abstimmung [Einstimmigkeit]. Nachdem Sie alle Nachrichten gesehen haben, erfolgt die Abstimmung. Sie können entweder für "rote Urne" oder "blaue Urne" stimmen. Ihre Stimme kann der Kugel, die Ihnen gezeigt wurde, oder der Nachricht, die sieht versendet haben, gleichen, muss aber nicht. Nur die Abstimmung zählt für Ihre Auszahlung.

Die Gruppenentscheidung muss einstimmig sein. Wenn alle Teilnehmer in Ihrer Gruppe für die "rote Urne" stimmen, ist die Gruppenentscheidung "rote Urne". Wenn alle für die "blaue Urne" stimmen, ist die Gruppenentscheidung "blaue Urne". Ansonsten beginnt eine zweite Abstimmungsrunde. Wenn nun alle drei Stimmen gleich sind, ergibt sich daraus die Gruppenentscheidung. Ansonsten gibt es eine dritte, finale Abstimmungsrunde. Wenn jetzt alle drei Stimmen gleich sind, ergibt sich daraus die Gruppenentscheidung. Andernfalls werden alle Stimmen, und damit die Gruppenentscheidung, auf rote Urne gestellt.

Fragebogen

1. Zuerst wählt der Computer eine Urne. Wie hoch ist die Wahrscheinlichkeit, dass der Computer die rote Urne wählt?

   ☐ 25% ☐ 50% ☐ 75%

2. Der Computer zeigt Ihnen eine Kugel, die zufällig aus der gewählten Urne gezogen wurde. Wenn die gewählte Urne blau ist, wie hoch ist die Wahrscheinlichkeit, dass Ihnen eine rote Kugel gezeigt wird?

   ☐ 40% ☐ 60% ☐ 80%

3. Richtig oder falsch?
   Nachdem Ihnen die Kugel gezeigt wurde, können Sie eine Nachricht versenden: rote Kugel oder blaue Kugel. Diese Nachricht muss mit der Ihnen gezeigten Kugel übereinstimmen.

4. Richtig oder falsch?
   Die Nachrichten aller drei Gruppenmitglieder werden allen Gruppenmitgliedern gezeigt. Danach können Sie abstimmen, und ihre Stimme darf nicht mit Ihrer Nachricht übereinstimmen.

5. Falls die gewählte Urne rot ist, Sie für die blaue Urne stimmten und die anderen beiden Teilnehmer für die rote Urne stimmten, wie hoch ist Ihre Auszahlung?

   ☐ 10 Taler ☐ 40 Taler ☐ 50 Taler

6. Falls die gewählte Urne blau ist, Sie für die rote Urne stimmten, ein anderer Teilnehmer für die rote Urne stimmte, und der dritte Teilnehmer für die blaue Urne stimmte, wie hoch ist Ihre Auszahlung?

   ☐ 10 Taler ☐ 40 Taler ☐ 50 Taler

7. Richtig oder falsch?
   Der Computer wird alle Teilnehmer zufällig in Gruppen einteilen, und in jeder Runde wird eine neue Einteilung vorgenommen.

C Experimental Instructions (translation)

This section contains a literal translation of both experimental instructions and control questionnaire (which originally are in German and available from the authors), and a composite screenshot displaying all the (German) words actually used in the experiment and their arrangement on the screen. This screenshot is composite in the sense that it displays all items at once (the message query, the vote query and the resulting payoff table) which in the experiment were displayed sequentially.

Instructions

This is an experiment in group decision making. Thank you for participating!

Please read these instructions very carefully. It is important that you do not talk to other participants during the entire experiment. In case you do not understand some parts of the experiment, please read through these instructions again. If you have further questions after hearing the instructions, please give us a sign by raising your hand out of your cubicle. We will then approach you in order to answer your questions personally. Please do not ask anything aloud.

The entire experiment will take place through computer terminals, and all interaction between you will take place through the computers. You will be paid for your participation in cash, at the end of the experiment. Different subjects may earn different amounts. What you earn depends partly on your decisions, partly on the decisions of others, and partly on chance.

The experiment consists of 50 rounds. The rules are the same for all rounds and for all participants. At the beginning of each round you will be randomly assigned to a group of 3 participants (including yourself). You will not know the identity of the other participants. After each round, groups will be randomly reassigned, but for simplicity we will always refer to the other participants in your group as "Participant 2" and "Participant 3". In each round you will only interact with the participants in your group. Your group will make a decision based on the votes of all group members. The decision is simply a choice between two jars, the blue jar and the red jar. In what follows we will explain to you the procedure in each round.

The Jar. There are two jars: the blue jar and the red jar. The blue jar contains 3 blue balls and 2 red balls. The red jar contains 3 red balls and 2 blue balls. At the beginning of each round, one of the two jars will be randomly selected. We will call this the selected jar. Each jar is equally likely to be selected, i.e. each jar is selected with a 50% chance. You will not be told which jar has been selected when making your decision.

The Ball. After a jar is selected for your group, the computer will show each of the participants in your group (including yourself) the color of one ball randomly drawn from that jar. Since you are three in your group, the computer performs this random draw three times. Each ball in the jar will be equally likely to be drawn for every member of the group. If the selected jar is blue, each member of your group has a chance of 60% of receiving a blue ball and a chance of 40% of receiving a red ball. If the selected jar is red, each member of your group has has a chance of 40% of receiving a blue ball and a chance of 60% of receiving a red ball. You will only see the color of your own ball.

The Message. After the ball has been presented to each of you, each player may send a message. The message is either "red ball" or "blue ball". The message you send may be equal to the ball you have been shown, or it may be different. It depends on your strategy and your preferences which message to send. When all group members have entered their messages, all of you will be shown all messages. Since there are three participants per group, each of you will see the same three messages, and each of these messages is either a red ball or a blue ball.

The Vote [Majority]. After all messages have been presented to each of you, each player is called to vote. You may vote either "red jar" or "blue jar". Your vote may but need not be related to the ball you have been shown or to the message you have sent. Only your vote and the group decision will affect your payoffs.

The group decision is determined by majority. If at least two participants in your group (including yourself) vote "red jar", then the group decision is "red jar". If at least two vote "blue jar", then the group decision is "blue jar".

The Vote [Unanimity]. After all messages have been presented to each of you, each player is called to vote. You may vote either "red jar" or "blue jar". Your vote may but need not be related to the ball you have been shown or to the message you have sent. Only your vote and the group decision will affect your payoffs.

The group decision has to be unanimous. If all participants in your group (including yourself) vote "red jar", then the group decision is "red jar". If all vote "blue jar", then the group decision is "blue jar". Otherwise, a second voting round starts. If all three votes are unanimous now, the decision is made. If it is again not unanimous, a third and final voting round starts. If all three votes are unanimous now, the decision is made. Otherwise, the group decision, and all individual votes, are set on red jar.

Payoff. Your payoff in each round is the sum of two components. First, if your group decision is equal to the correct jar, each member of your group earns 40 Talers. If your group decision is incorrect, each member of your group earns 0 Talers from the group decision. Second, if your individual vote is "red jar", you earn an additional 10 Talers. If your individual vote is "blue jar", your additional payoff is 0 Talers. Depending on which jar had been selected by the computer, the following tables summarize the possible outcomes.

The computer selected the **blue jar**

	The other two votes are			
Your vote	Blue + Blue	Blue + Red	Red + Blue	Red + Red
Blue jar	40	40	40	0
Red jar	50	10	10	10

The computer selected the **red jar**

	The other two votes are			
Your vote	Blue + Blue	Blue + Red	Red + Blue	Red + Red
Blue jar	0	0	0	40
Red jar	10	50	50	50

Information at the end of each Round. Once you and all the other participants have voted, the round will be over. At the end of each round, you will receive the following information about the round: messages and votes of all players, the group decision, the jar selected by the computer, your payoff.

Final Earnings. At the end of the experiment, the Talers earned in all 50 rounds are added up and converted to Euro. Each Taler is converted to 1 Cent. Thus, 100 Talers are converted to 1 Euro. Additionally, you will earn a show-up fee of 5.00 Euros. Everyone will be paid in private and you are under no obligation to tell others how much you earned.

Questionnaire (computerized)

1. What is the probability that the computer selects the red jar?

   ☐ 25% ☐ 50% ☐ 75%

2. The computer shows you exactly one ball drawn randomly from the selected jar. If the selected jar is blue, what is the probability that you are shown a red ball?

   ☐ 40% ☐ 60% ☐ 80%

3. Right or wrong?
   After having been shown the ball, you can send a message: red ball or blue ball. This has message has to be equal to the ball you have been shown.

4. Right or wrong?
   The messages of all three group members are shown to all group members. Subsequently, you can vote, and the vote must be different from the message you have sent.

5. If the selected jar is red, you voted "blue jar" and the other two players voted "red jar", what is your payoff?

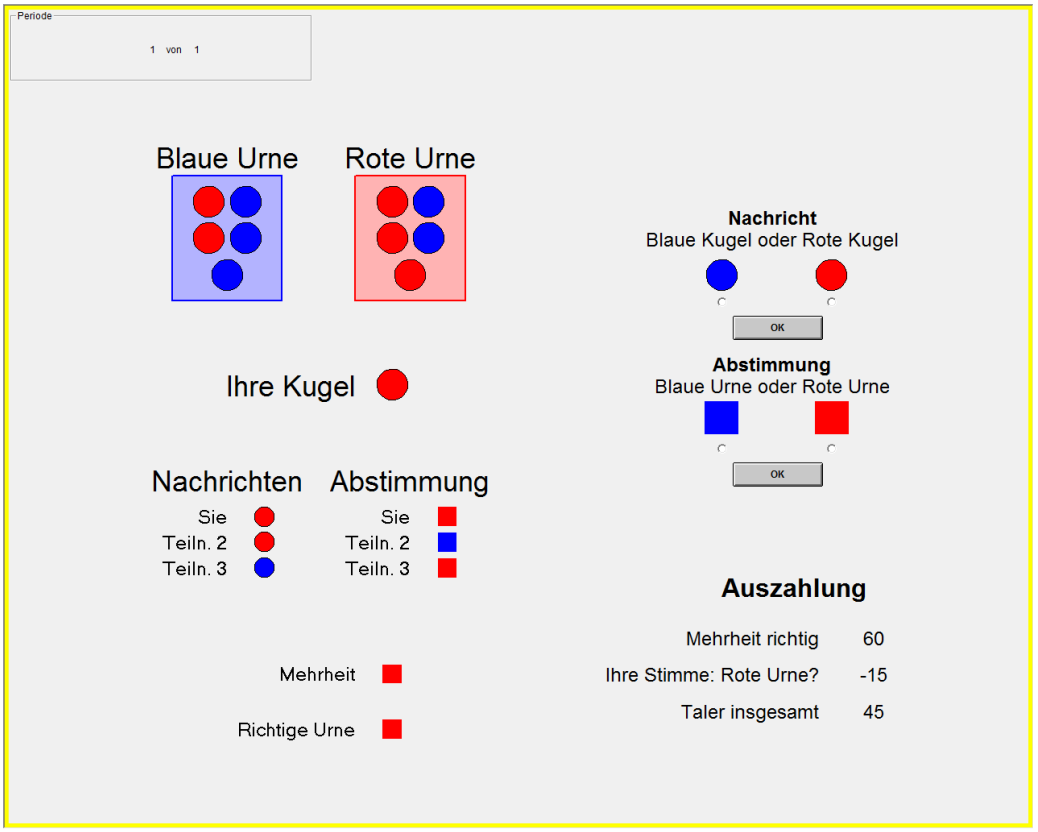
☐ 10 Taler ☐ 40 Taler ☐ 50 Taler

6. If the selected jar is blue, you voted “red jar”, one other player voted “red jar”, the third one voted “blue jar”, what is your payoff?

   ☐ 10 Taler ☐ 40 Taler ☐ 50 Taler

7. Right or wrong?
   The computer will assign all participants randomly to groups, and in each round, a new random assignment will be made.

Figure 2: Composite screenshot (in German)



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 drawn ball (entire game)
   Shows the two jars ("Blaue Urne" and "Rote Urne" means "blue jar" and "red jar") and the ball drawn ("Ihre Kugel" means "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 "Nachricht – Blaue Kugel oder Rote Kugel" (Mesage – Blue Ball or Red Ball) 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 (for rest of game)
   Now the box "Nachrichten" (Messages) on the left appears, with the messages of all three subjects. "Sie" means "You", "Teiln. 2" means "Co-Participant 2", and "Teiln. 3" means "Co-Participant 3". 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 "Abstimmung – Blaue Urne oder Rote Urne" (Vote – Blue Jar or Red Jar) appears with the two rectangular jars underneath 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 (for rest of game)
   Now the box "Abstimmung" (Votes) on the left appears, with the votes of all three subjects. "Sie" means "You", "Teiln. 2" means "Co-Participant 2", and "Teiln. 3" means "Co-Participant 3". 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 voting stage is restarted).
6. After five seconds, the decision taken by the committee ("Mehrheit" means majority), the true jar chosen by Nature ("Richtige Urne" means true jar) and the payoff information is displayed. "Auszahlung" means payoff, "Mehrheit richtig" means "majority correct", "Ihre Stimme: Rote Urne?" means "Your Vote: Red Jar?", and "Taler insgesamt" means "Taler in total" (where "Taler" is our experimental currency unit). This information remains on the screen for 10 seconds. Note that voting "Red" in this screenshot is associated with minus 15 Taler for testing purposes, the payoffs used in the experiment were plus 10 or plus 15, as described in the paper.