Breitmoser, Yves; Valasek, Justin

Working Paper

Why Do Committees Work?

CESifo Working Paper, No. 10800

Provided in Cooperation with:
Ifo Institute – Leibniz Institute for Economic Research at the University of Munich

Suggested Citation: Breitmoser, Yves; Valasek, Justin (2023) : Why Do Committees Work?, CESifo Working Paper, No. 10800, Center for Economic Studies and ifo Institute (CESifo), Munich

This Version is available at:
https://hdl.handle.net/10419/282488

CESifo WORKING PAPERS

10800
2023

November 2023

Why Do Committees Work?

Yves Breitmoser, Justin Valasek

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 10800*

Why Do Committees Work?

Abstract

We report on the results of an experiment designed to disentangle behavioral biases in information aggregation of committees. Subjects get private signals about the state of world, send binary messages, and finally vote under either majority or unanimity rules. Committee decisions are significantly more efficient than predicted by Bayesian equilibrium even with lying aversion. Messages are truthful, subjects correctly anticipate the truthfulness (contradicting limited depth of reasoning), but strikingly overestimate their pivotality when voting (contradicting plain lying aversion). That is, committees are efficient because members message truthfully and vote non-strategically. We show that all facets of behavior are predicted by overreaction, subjects overshooting in Bayesian updating, which implies that subjects exaggerate the importance of truthful messages and sincere voting. A simple one-parameteric generalization of quantal response equilibrium capturing overreaction covers 87 percent of observed noise.

JEL-Codes: D710, D720, C900.

Keywords: committees, incomplete information, cheap talk, information aggregation, laboratory experiment, Bayesian updating, lying aversion, limited depth of reasoning.

*Yves Breitmoser**
University of Bielefeld / Germany
yves.breitmoser@uni-bielefeld.de

Justin Valasek
Norwegian School of Economics
Bergen / Norway
justin.valasek@nhh.no

*corresponding author

November 21, 2023
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.

1 Introduction

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

While committees are a ubiquitous institution for making decisions, they are often derided as ineffective and incompetent. In contrast to popular opinion, however, the experimental evidence has shown that relative to theoretical benchmarks, committees work surprisingly well. In particular, committee members are less strategic, more truthful, and are ultimately better at aggregating individual information than we should expect.[1] Despite this evidence, to the best of our knowledge there is no systematic study of why, exactly, committees work better than expected.

A basic theoretical argument for the use of committees, first formalized by de Condorcet (1785), is that committees aggregate the private information of their members and thus make more informed decisions than could be made by any one individual in isolation. If all individuals hold private information that is more likely to be "right" than "wrong," and if all individuals vote according to their private information, then a sufficiently large committee that votes via a majority rule will choose the "right" option with arbitrary precision. However, starting with Austen-Smith and Banks (1996) and Feddersen and Pesendorfer (1996, 1997) a large theoretical literature emerged showing that, in general, committee members do not always have incentives to vote sincerely, and that the prediction of information aggregation in committees is not robust to many real-world concerns.[2] Additionally, while truthful communication may mitigate some of the failures of information aggregation through voting (Coughlan, 2000), committee members often face incentives to communicate strategically and misrepresent their individual information (Austen-Smith and Feddersen, 2006; Breit-

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

[2] Theoretically, the prediction of information aggregation in committees can fail due to: the decision rule (Feddersen and Pesendorfer, 1998); uncertainty regarding the signal structure (Mandler, 2012); a failure of preference monotonicity (Bhattacharya, 2013); uncertainty regarding the size of the population (Ekmekci and Lauermann, 2019); and voters who are bribed (Dal Bó, 2007), who have a preference for winning (Callander, 2007, 2008), who have moral payoffs (Feddersen et al., 2009), who have partisan expressive motives (Morgan and Várdy, 2012), and are held accountable for their individual vote (Midjord et al., 2017, 2021).

moser and Valasek, 2022).

This raises a simple question given the empirical evidence on committees: Do we misperceive the efficiency of committees empirically, or are committees more efficient than predicted theoretically? Despite the large amount of work dedicated to analyzing committees theoretically and experimentally, we are not aware of prior work directly answering this simple question, much less of a study answering it for committees allowing for communication. We present a simple experimental design, providing committee members with incentives to both communicate and vote strategically, that allows us to answer this question. The design is simple in that we implement standard voting games with (formalized) cheap-talk communication, but it is sufficiently rich in that we allow for expressive payoffs (that agents have a preference to vote for one of the options, Breitmoser and Valasek, 2022) under both majority and unanimity voting. This set-up generates interior equilibria with strategic communication and voting, in relation to which both over- and undercommunication as well as naive and sophisticated voting is possible. We find that committees indeed aggregate information more efficiently than predicted theoretically, largely in line with socially optimal behavior: subjects tend to communicate truthfully and vote with the majority of messages (see also Le Quement and Marcin, 2020). This finding raises another question, however: Why do committee members behave in this fortunate way?

On one hand, our analysis reveals that subjects mostly understand that all of them tend to communicate truthfully, but in light of this, the expressive payoffs in our experiment should induce them to vote strategically. The correct anticipation of truthful communication suggests that limited depth of reasoning is not the key driver of behavior, as we discuss in more detail below. On the other hand, the truthful communication itself has been linked to various models in the literature, most prominently lying aversion, guilt aversion and limited depth of reasoning. While the evidence between the models in the existing literature is generally inconclusive, as discussed shortly, in our experiment, subjects would nonetheless have an incentive to vote strategically by collecting the expressive payoff. Our empirical analysis falsifies this prediction, however, implying that the behavioral foundation of the

observed truthful communication and sincere voting follows from a different behavioral bias.

Before we elaborate, let us clarify the exact timing of committee decisions in our model: Nature draws a state of the world, agents receive an informative signal about the state and update their belief, agents then send messages and again update their belief, and finally they vote. In relation to this framework, our central empirical observations can be summarized as follows: (i) subjects communicate much more truthfully than justified by expected payoffs, (ii) they correctly anticipate the imperfect truthfulness of their co-players' messages, and (iii) they vastly overweigh private signals and public messages in the voting stage. Observations (i) and (iii) contradict plain Bayesian equilibrium, observation (iii) contradicts Bayesian equilibrium with lying aversion, and observation (ii) contradicts limited depth of reasoning.

Based on our understanding of human behavior, however, there are two obvious breaking points in the above chain of events: we are not good at Bayesian updating of our beliefs. Taking this as given, our behavioral analysis asks how much of behavior is unexplained once we allow for the one factor that is inevitably at play—imperfect Bayesian updating. The answer turns out to be "not much," as seen in our simple but general structural analysis nesting the three debated biases (lying aversion, limited depth of reasoning and overreaction) besides a number of other biases such as efficiency concerns and limited foresight. All three main biases affect behavior significantly, but to varying extents, and by far the most substantial factor driving behavior is overreaction: While Bayesian equilibrium (with logistic errors) explains around 40% of observed behavior, allowing for imperfect Bayesian updating explains another 50%, yielding 87% in total. That is, once we account for imperfect Bayesian updating, other factors are of limited relevance.

To summarize, we observe behavior compatible with *overreaction* to the most recent information (see Bordalo et al., 2020; Afrouzi et al., 2023), which in our setting translates to a tendency for subjects to overestimate the probability that the state equates with the private signal(s) . Such overreaction strengthens incentives to message truthfully, and as subjects overestimate the relevance of their own decisions, it also strengthens incentives

to vote with the majority. Moreover, it is difficult to learn correct Bayesian updating in an applied setting, explaining its persistence over the course of the experiment and, as we discuss in our conclusion, it is a general explanation for many of the related observations made in prior experiments.

Regarding related literature, in a companion paper (Breitmoser and Valasek, 2022) we discuss the policy implications of our experimental results for the choice between majority and unanimity rules, while the present paper discusses the more fundamental question about the (exceeding) efficiency of committees. Our experimental design builds on Goeree and Yariv (2011), who analyze voting in committees with pre-play communication and similarly find overcommunication. Their work extends Guarnaschelli et al. (2000) and Ali et al. (2008), who studied voting in committees without pre-play communication. In relation to this literature, our experiment adds expressive payoffs which greatly facilitates the identification of behavioral factors through implying distinctive predictions of the various concepts discussed in the literature.

As indicated, the behavioral factors discussed in the existing literature are primarily lying aversion, guilt aversion and limited depth of reasoning. To begin with, the evidence between lying aversion and guilt aversion is rather mixed. While the original studies of Gneezy (2005) and Charness and Dufwenberg (2006) provide rather strong evidence in favor of lying aversion and guilt aversion, respectively, subsequent studies cast doubt either on the validity of guilt aversion (Vanberg, 2008; Ellingsen et al., 2010), lying aversion (Hurkens and Kartik, 2009), or both (Charness and Dufwenberg, 2010, 2011). Overreaction, recently highlighted as a simple behavioral deviation from Bayesian updating in Bordalo et al. (2020) and Afrouzi et al. (2023), has the potential to be an organizing factor in much of this work, as imperfect Bayesian updating is a widespread phenomenon of obvious relevance in communication in the presence of private information.

Another recent paper discussing policy implications of expressive payoffs is Ginzburg et al. (2022), who observe limited depth of reasoning, which we do not find to be a major factor after allowing for overreaction. A seminal paper demonstrating that limited depth of

reasoning is not necessarily driving truthful communication is Cai and Wang (2006), who find that behavior is well-explained by logit QRE (note that we will also allow for logistic errors). Sánchez-Pagés and Vorsatz (2007) find that some subjects additionally are lying averse (while holding rational expectations about the noisy play), and further evidence indicates that subjects expect more truthfulness than predicted in equilibrium (Gneezy, 2005; Kawagoe and Takizawa, 2008; Lundquist et al., 2009). This phenomenon is known as truth bias, whereby receivers rely more on the senders' messages in choosing actions than predicted by equilibrium. In our set-up with interior equilibria, receivers actually rely more on the senders' messages in choosing actions than predicted by rational expectations, which ultimately points to overreaction as the behavioral bias with the potential of organizing this large array of observations.

We shall discuss this potential in more detail in our concluding section 5. Aside from this, Section 2 discusses the theoretical model and the experimental design, section 3 analyzes the rationality of communication and voting in isolation, while section 4 analyzes both in conjunction to quantify the relevance of the key behavioral factors. The appendix contains further information on our theoretical analysis and the experimental instructions.

2 Theory and Experimental Design

2.1 The voting game

We consider a generalized Condorcet setting with pre-vote communication (cheap talk) and expressive payoffs: First, Nature draws a state of world $\omega \in \{R(ed), B(lue)\}$, and $N = 3$ experts, $i \in \{1,2,3\}$, draw independent signals about the state of the world. That is, each committee member receives a private signal, $s_i \in \{R,B\}$, which are identically and independently distributed across members with $\Pr(\omega = x | s_i = x) = \alpha$. We set α equal to 0.6 in the experiment.

Each expert then sends a message $m_i \in \{R,B\}$ to the others and finally the committee

decision, denoted by $X \in \{R,B\}$, is made via a vote, where each expert submits a vote, $v_i \in \{R,B\}$, simultaneously with no abstentions. The voting rule, D, is either Majority or Unanimity (defined below) and maps votes (v_i) to committee decisions (X). All committee members have a common prior over the state of the world, denoted $P_R = \Pr(\omega = R)$, which is uninformative, i.e. $P_R = 1/2$.

The payoff of expert i is a function of the committee decision, the state of the world $\omega \in \{R,B\}$, and the own vote v_i. Payoffs consist of a common-value component and an idiosyncratic component. For the common-value component, each candidate receives a payoff of P_c if the committee chooses the decision that matches the underlying state of the world, and a payoff of zero otherwise. For the idiosyncratic component, each candidate has a *voting bias*, and receives a payoff of $K < P_c$ conditional on voting for option R. This voting bias is unconditional on the state of the world and the decision of the committee – such an idiosyncratic payoff is commonly referred to as an expressive payoff (Breitmoser and Valasek, 2022; Ginzburg et al., 2022). To summarize, terminal payoffs as a function of outcome $X \in \{R,B\}$, state $\omega \in \{R,B\}$, and the own vote $v_i \in \{R,B\}$ are equal to:

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

Crucially, the addition of idiosyncratic payoffs allows us to distinguish between the theoretical predictions of different behavioral models. As we will detail in the theoretical predictions below, an expressive payoff introduces a collective-action problem that gives an incentive to communicate strategically. The main behavioral models we consider, lying aversion and overreaction, both predict more truthful communication relative to the benchmark, but give different predictions in the voting stage.

We denote this game by $\Gamma = \langle K, P_R, \alpha, D, N \rangle$. The timing of the game is as follows:

1. Nature draws state $w \in \{R, B\}$ and sends private signals $(s_i) \in \{R, B\}^N$.
2. Committee members observe s_i and simultaneously send messages $(m_i) \in \{R, B\}^N$.
3. Committee members observe (m_i) and simultaneously submit votes $(v_i) \in \{R, B\}^N$.
4. Votes are counted and payoffs accrue.

The Majority decision rule is defined in the standard way: $X = B$ iff two or more agents vote for B, and $X = R$ otherwise. For Unanimity, we allow for multiple rounds for the committee to reach a unanimous decision in our experiment.[3] To simplify the exposition of the theoretical results, however, we use a model of unanimity that mechanically enforces a unanimous voting profile if agents do not reach a unanimous decision in the first round, since the additional rounds of straw polls do not impact our theoretical results (see Breitmoser and Valasek, 2022). That is: $X = B$ iff $v_i = B$ for all i; if $v_i = R$ for any i, $X = R$ *and* $v_i = R$ for all i.

We focus on symmetric strategies, so that agents' strategies are duples (σ, τ), where:

- $\sigma(s_i)$ is the probability of message R after signal $s_i \in (R, B)$,
- $\tau(s_i, m_i, M)$ is the probability of vote $v_i = R$ after signal s_i, own message m_i, and overall $M = \#\{j | m_j = B\}$ players sending message B (including i).

We consider several different models of behavior, and therefore introduce the associated equilibrium concepts in the following section.

2.2 Theoretical predictions

In this section we introduce several behavioral models, and summarize the theoretical predictions of the various models under the parameters used in the experiment: a precision, α, equal to 0.6; and two levels of expressive/common payoffs, "low" expressive payoffs ($K = 10$, $P_c = 40$), and "high" expressive payoffs ($K = 15$, $P_c = 35$). Our benchmark solution concept is efficient symmetric sequential equilibrium, or "equilibrium" for short. In

[3]This approximates the general process often used in committees (see Guarnaschelli et al., 2000; Goeree and Yariv, 2011): agents first publicly share their private signals (cheap talk messaging), then "deliberate" using a sequence of straw polls, and reach a final decision when the outcome of a straw poll is unanimous

addition, we will provide theoretical predictions for several behavioral models to clarify how our experimental design allows us to discriminate the main strands of ideas potentially underlying committee behavior: non-standard preferences (lying aversion and efficiency concerns), non-Bayesian updating (overreaction), and non-equilibrium beliefs (of which we focus on level-k).

Next, we will outline the main theoretical findings and brief intuition for the observed differences. All theoretical predictions are also summarized in Table 15 in the Appendix. Since there are often multiple equilibria in our setting, we focus on efficient equilibria, i.e. the equilibrium that maximizes expected payoffs. If there are multiple efficient equilibria, we focus on the equilibrium that is the limiting logit equilibrium, which is the limiting agent quantal response equilibrium (McKelvey and Palfrey, 1998) for logit errors as the error rate tends to zero. As McKelvey and Palfrey demonstrate, this limiting equilibrium is a sequential equilibrium, which connects to the existing theoretical literature, and as we shall consider logit errors below, the limiting logit equilibrium also connects to our structural analysis of behavior reported below.

The following table (Table 1) summarizes the predictions of our main behavioral models under Majority, indicating whether subjects send truthful messages, and whether the committee votes with the majority of messages. The main differences in predictions between our benchmark equilibrium and the behavioral equilibria are highlighted, showing that there are substantial differences in the predictions across the different models of behavior. Note that, while illustrative, the coarse predictions in Table 1 obscure some substantial variation; for example, while both lying aversion and level-L (to be defined) predict that the committee votes with the majority of messages when $M = 3$ for low expressive payoffs, the individual probability of $v_i = B$ is 0.64 under lying aversion, while the individual probability of $v_i = B$ is 1 under level-L (see Table 15 in the Appendix for additional detail).

Benchmark Equilibrium As indicated, the baseline equilibrium concept we consider is symmetric sequential equilibrium. By symmetry, we mean that agents with the same infor-

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 *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."

mation sets take the same strategies.

Definition 1. *Our benchmark equilibrium (equilibrium) is symmetric sequential equilibrium. In cases of multiple equilibria, we focus on the equilibrium that maximizes ex ante expected utility.*

In Breitmoser and Valasek (2022), we characterize the set of equilibria for the parameters we use in our experiment in greater detail. In particular, we show that under Majority, the probability that the committee selects $X = B$ when there are three signals for B is strictly smaller than 1 in all equilibria. For intuition, assume agents communicate truthfully and vote for the state supported by the majority of messages. Under Majority, the players are not pivotal given this voting behavior, and therefore have an incentive to unilaterally deviate to vote R whenever B has the higher number of messages (and signals). This shows that voting with the majority of signals cannot be an equilibrium. This "collective action" problem prevents the committee from selecting the optimal committee outcome given the profile of messages and causes strategic communication in any equilibrium where agents' voting strategies respond to the aggregate profile of messages.

Prediction 1 (Equilibrium: Majority)**.** *For low expressive payoffs,* $\sigma(R) = 1$*,* $\sigma(B) = 0.56$ *and* $v_i = B$ *iff* $s_i = B$ *and* $M = 2$*. For high expressive payoffs,* $\sigma(R) = 0.5$*,* $\sigma(B) = 0.5$ *(babbling) and* $v_i = R$ *for all message profiles.*

That is, for high expressive payoffs, the collective action problem causes information aggregation to collapse completely, in the sense that the committee selects $X = R$ for all message profiles. For low expressive payoffs, the committee partially aggregates information. In particular, when there are two messages for B, agents with $m_i = B$ vote for B and the agent with $m_i = R$ votes for R, which implies that agents voting for B are pivotal. In turn, it is a best reply for agents to vote for B when pivotal when $M = 2$ because the probability that there are three signals for B when $M = B$ is strictly positive given $\mu(B) < 0$ (which is not the case under truthful communication).

Lying aversion As discussed above, the evidence from previous experiments on communication in committees shows that communication is more truthful than predicted in equilibrium—a finding that we replicate in our setting. A potential explanation for this finding is that agents are averse to lying (Gneezy, 2005; see Abeler et al., 2019 for evidence on lying aversion). We therefore consider a simple model of lying aversion, where agents play truthful messaging strategies.

Definition 2 (Lying aversion)**.** *We define lying aversion as agents playing a fixed messaging strategy,* $\bar{\sigma} = \{\sigma(a) = 1, \sigma(b) = 0\}$.

The model under lying aversion reduces to a game where all agent's signals are publicly shared, which results in the following voting behavior under majority:

Prediction 2 (Lying Aversion: Majority)**.** *By definition,* $\sigma(R) = 1$*,* $\sigma(B) = 0$*. For low expressive payoffs, agents only vote for B if* $M = 3$ *(*$\Pr(v_i = B|M = 3) = 0.64$*). For high expressive payoffs,* $v_i = R$ *for all message profiles.*

Note that lying aversion results in truthful messaging by assumption, but does not address the collective action problem in the voting stage. Therefore, since agents report truthfully

and voting B is only individually rational given three signals for B, agents will only vote for B with positive probability when expressive payoffs are low, and there are three messages for B.

Overreaction Players exhibit overreaction if they tend to ignore prior information as new information arrives (see Bordalo et al., 2020; Afrouzi et al., 2023). For our theoretical benchmark, we implement (perfect) overreaction as the tendency to assume that the most recent information is correct with high probability, regardless of the prior information about the state of the world and the probabilities of signals in each state.

Definition 3 (Overreaction)**.** *We define overreaction as agents perceiving private signals to be correct with a probability close to 1—i.e.* $\Pr(\omega = s_i) \to 1$*—and where this is common knowledge.*

Our model of overreaction presumes that agents are "rational," in the sense that they choose strategies to maximize their individual payoffs and play according to the benchmark equilibrium concept, but that their beliefs are not consistent with Bayesian updating. Under Majority, similar to lying aversion, incorrect Bayesian updating leads to more truthful communication relative to the benchmark. The reason is that since players with overreaction overestimate the probability that their own signal is correct, these players underestimate the value of sending a strategic message. That is, conditional on receiving a message of R, the strategic incentive to report B to increase the probability that B is chosen when the other agents received signals of B is outweighed by the chance that the committee incorrectly chooses B when the other agents' signals are split.

Prediction 3 (Overreaction: Majority)**.** *With Overreaction, agents truthfully communicate* $\sigma(R) = 1$, $\sigma(B) = 0$. *For* $M = 2$ *agents with* $m_i = B$ *set* $v_i = B$ *with both high and low expressive payoffs, and for* $M = 3$ *all agents set* $v_i = B$ *with probability* 0.85 *for low expressive payoffs and* 0.69 *with high expressive payoffs.*

That is, in contrast to lying aversion, overreaction impacts voting relative to the benchmark behavior. Since agents anticipate that messaging will be truthful, but overestimate the

accuracy of the signals, they will vote for B with positive probability given two messages of B. In contrast, with lying aversion, agents only vote for B given three messages of B. Therefore, while lying aversion and overreaction both predict more truthful messaging relative to the benchmark, only overreaction predicts that agents will vote according to the majority of messages at a higher rate than the benchmark in equilibrium.

Limited depth of reasoning Here we consider two models of limited depth of reasoning. First, adopting the standard assumption that players at level 0 randomize uniformly and that players at level $k \geq 1$ best respond to players at level $k-1$, yields the following prediction of level-1 behavior.

Definition 4 (Level-K)**.** *Under $k = 1$, agents best response to the belief that all opponents randomize uniformly in both the communication and voting stages.*

Given uninformative communication and randomized voting, agents have a best response of voting for R for all profiles of signals. The predictions of cursed equilibrium, specifically for fully cursed players, are very similar, and for this reason, we shall skip a detailed discussion. Due to their similarities in our setting, we will jointly refer to these models as models of limited depth of reasoning.

Arguably, however, the idea that level-K agents believe that others will randomize uniformly in the communication stage is implausible in our context, considering our knowledge of lying aversion. For this reason, we also consider the alternative model of limited depth of reasoning where agents are level-K with infinitesimal lying aversion: in cases of indifference agents will communicate truthfully (including agents at level 0), and level 0 agents will vote by randomizing uniformly. We shall refer to this combination of level-K and lying aversion as the level-L model.

Definition 5 (Level-L)**.** *Under level $l = 1$, agents best response to the belief that all opponents communicate truthfully and randomize uniformly in the voting stage.*

The predictions of the model in the voting stage change with Level-L relative to Level-K

since, given three signals for B, agents best respond to randomized voting by voting B with low expressive payoffs.

Altruism Another point of view is that players might be (imperfect) altruists, and place a higher weight on the common payoffs (see for example Ginzburg et al., 2022).[4]

Definition 6 (Altruism)**.** *We define "Altruism" as agents disregarding the expressive payoffs, and playing strategies that maximize common payoffs.*

The predictions under altruism are straightforward: agents babble in the communication stage, but vote according to their signal ($v_i = s_i$), ensuring that the option with the majority of signals is selected by the committee.

2.2.1 Unanimity

Here we briefly address the predictions under Unanimity, summarized in Table 2 (note that the table list voting behavior as a function of *signals* rather than of messages as in Table 1). In contrast to Majority, under Unanimity there is no collective action problem—the committee will select B if and only if all agents vote for B. In general, this implies that there are many voting strategies that result in the same committee outcome, and there are multiple efficient equilibria for many of our models. Again, when there are multiple efficient equilibria, we focus on the limit logit equilibrium.

In our benchmark model, the committee will only select $X = B$ when there are three signals for B using a messaging strategy of babbling, $\mu(A) = \mu(B) = 0.5$, and a voting strategy of $v_i = s_i$. Under lying aversion, the committee will also only select $X = B$ when there are three signals for B, but will play truthful strategies in the messaging stage.

Under overreaction, agents communicate truthfully. However, they vote for the option that is optimal given their beliefs, which implies that agents will vote for the option with the majority of messages. That is, the committee will select $X = B$ if there are two or more

[4] Alternatively, agents may derive utility from voting for the correct alternative (see Midjord et al., 2021).

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."

messages for B. Similar to under Majority, overreaction predicts that the committee will select B more often than the benchmark equilibrium or lying aversion (the same is true for altruism).

For level-K, the predictions are the same for Majority—agents babble in the messaging stage and vote for R for all message profiles. Under level-L, however, the committee selects B iff there are three signals for B since agents communicate truthfully and vote for B given three messages for B.

2.3 Experimental design

The experiment closely implements our model of voting with expressive payoffs, using a 2×2 design with "High" and "Low" expressive payoffs under Majority and Unanimity. The experimental implementation closely follows the related experiments of Guarnaschelli et al. (2000) and Goeree and Yariv (2011). In particular, we use neutral language, communicate probabilities and signals to subjects using balls drawn from urns, and provide feedback about

the actual state of world and composition of payoffs after each round. A detailed description follows and a translation of the instructions and a screenshot are provided as supplementary material. The experiments were conducted at the WZB/TU experimental laboratory in Berlin in May, June and November of 2016. Subjects were recruited using ORSEE (Greiner, 2015) and the experiment was programed in Z-Tree (Fischbacher, 2007).

The four treatments are summarized in Table 3. The sum of the expressive payoff K, which committee members get after voting R, and social payoff P_c, which committee members get after voting in line with the state of the world, is always equal to to 50 points. In the treatments with "low" expressive payoffs, we set $K = 10$ and $P_c = 40$, and in the treatments with "high" expressive payoffs, we set $K = 15$ and $P_c = 35$. The payoffs were calibrated such that in both the Low and High treatments, voting for B is not individually rational when there are only two signals for B (as discussed above). For both cases of expressive payoffs, we conduct sessions with either Unanimity or Majority voting, implementing all treatments strictly between subjects. The precision of each subject's signal was constant across treatments and equal to $\alpha = 0.6$. Subjects were paid according to the sum of points accumulated across all 50 games, and one point corresponded to one euro cent in all treatments. The experiment lasted between 75 and 105 minutes and subjects earned between 19 and 22 Euros on average across sessions.

Table 3: Overview of experimental treatments

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

For each treatment, we ran four sessions with either 9 (two sessions) or 12 (fourteen sessions) participants. In all cases, two sessions were run simultaneously to increase anonymity. Upon arrival at the laboratory, subjects were seated randomly. An experimental assistant then handed out printed versions of the instructions and read the instructions out loud. Subsequently, subjects filled in a computerized control questionnaire verifying their understand-

ing of the instructions, and the experiment did not start until all subjects had answered all questions correctly. The subjects then played 50 voting games in committees of size three ($N = 3$), with random rematching after each game (see Figure 2 in Appendix B for a composite screenshot). After each game, the subjects received feedback on payoffs, aggregate behavior and the aggregate signal profile. Under Majority, the timing of each round was as follows.

Definition 7 (Implementation: Majority voting)**.** *Observe private signal $s_i \in \{R,B\}$. Send a public message to their group $m_i \in \{R,B\}$. Observe message profile. Submit vote $v_i \in \{R,B\}$. Observe state, votes, outcome, and payoffs for this game.*

Under Unanimity, the timing of each round was identical to Majority aside from the voting stage. Subjects were given three chances to reach a unanimous decision—again, the additional rounds of "straw polls" do not impact our theoretical predictions—after which all subjects were assigned a default vote of R.

Definition 8 (Implementation: Unanimity voting)**.** *The only difference to Majority is in the voting stage: Submit vote $v_i^1 \in \{R,B\}$. If vote is unanimous proceed to Outcome Stage. Otherwise, again submit a vote $v_i^2 \in \{R,B\}$. If vote is unanimous proceed to Outcome Stage. Otherwise, submit a final vote $v_i^3 \in \{R,B\}$. If vote is not unanimous, all subjects are assigned the vote $v_i = R$. Proceed to Outcome stage.*

Upon completion of the experiment, subjects left the laboratory and were paid individually in a separate room by an experimental assistant.

3 Are committees as efficient as predicted?

Table 4 displays the average strategies across treatments and Table 5 provides the implied efficiency of information aggregation, i.e. the relative frequencies that the committees choose the options corresponding with the majority of their private signals. Theoretically, efficient information aggregation is achieved if, under majority, subjects simply vote according to

Table 4: Observed choices in the experiment

	Messages		Voting											
	$\mu(A)$	$\mu(B)$	$\pi(A,A,0)$	$\pi(B,A,0)$	$\pi(A,A,1)$	$\pi(A,B,1)$	$\pi(B,B,1)$	$\pi(B,A,1)$	$\pi(A,A,2)$	$\pi(A,B,2)$	$\pi(B,B,2)$	$\pi(B,A,2)$	$\pi(A,B,3)$	$\pi(B,B,3)$
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
Base Rate Fallacy	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.

their signals, and under unanimity, if subjects message truthfully and vote according to the majority of messages. As discussed above, subjects may want to deviate from voting efficiently to collect the expressive payoff, and in this respect, truthful messaging may be detrimental: The more subjects know about their co-players' signals, the more they can infer about their subsequent voting behavior and the easier it is to identify opportunities for collecting the expressive payoffs. Indeed, theoretically lying aversion implies more truthful messages than Bayesian equilibrium for expected payoffs but weakly less efficient committee decisions. Overall, socially efficient Bayesian equilibrium is the most optimistic prediction and predicts efficiency rates 0.617, 0.5, 0.64 and 0.64 across treatments (Table 5).

The observed efficiency rates, across treatments and for both inexperienced and experienced subjects (first and second halves of sessions, respectively), are substantially higher than this prediction: 0.681, 0.669, 0.777 and 0.732. Indeed, taking each of the sixteen sessions as independent observations, the observed efficiency rate is below the most optimistic equilibrium prediction in only one half of one session: the subjects in the second half of session 4 on the 35-15 unanimity treatment. One out of 2×16 half sessions is clearly within the limits of chance for any statistical test, and for example, using Wilcoxon tests, the Null hypothesis that efficiency rates do not exceed the most optimistic Bayesian equilibrium predictions is rejected at $p < 10^{-5}$. That is, committees do work better than predicted by the models discussed prominently in the existing literature (equilibrium, lying aversion, and limited depth of reasoning).

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

While this was our initial hypothesis, let us note that the only concepts predicting more efficient information aggregation than Bayesian equilibrium are overreaction and altruism, of which only overreaction predicts an asymmetry between unanimity and majority voting. Table 5 provides the predictions for full overreaction, i.e. that subjects believe their signal is correct with a probability close to 1 (which is true for all players and common knowledge). This predicts much more efficient information aggregation than we actually observe, but it

seems that the deviations from Bayesian equilibrium point to this direction, including the asymmetry, e.g. that subjects potentially exhibit overreaction to an intermediate degree. In the remainder of the paper, we shall analyze potential explanations in detail, starting by looking at messages and votes in isolation.

4 Do subjects communicate rationally?

We begin with analyzing communication behavior in relation to the actually expected payoffs associated with the messages across information sets in the experiment. To this end, we determine the expected payoffs of subjects in the experiment, conditional on the signal received and message sent (Table 6). The right-most columns ("Rel Freq *R*") in the panels of Table 6 provide the observed relative frequencies of *R* messages. Following an *R* signal, they are largely constant around 0.9, and following a *B* signal and they are also largely constant around 0.15, implying that messages are slightly more truthful in unanimity treatments. Messages appear to be largely independent of the expected payoffs associated with messaging *R* or *B*, however, which are displayed in the other two columns. The probabilities of *R* messages following *B* signals are always around 0.15, but expected payoffs vary from favoring *B* message (majority 40-10) over indifference (unanimity 40-10) to favoring *R* messages (35-15 treatments). Similarly, following *R* signals, by expected payoffs subjects in both majority treatments should be practically indifferent between *R* and *B* messages, but the probability of *R* messages is always high, and without obvious reason, it drops from 0.904 to 0.787 between these treatments. In unanimity treatments, subjects are more truthful following *R* signals than following *B* signals, and indeed, incentives are strongly in favor of *R* messages in these cases. In isolation, the latter may relate to expected payoffs, but the overall impact of expected payoffs (under rational expectations) seems limited in communication.

In order assess this relationship rigorously, we regress the probability of messaging $m = R$ on signal and expected payoffs in the ensuing voting stage. Formally, we use $dEP(s)$ to denote the difference in expected payoffs between messages *R* and *B* given a subject's signal

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: This tables provides the conditional payoffs observed in the experiment, conditional on the signal being *R* or *B* and the subject sending either an *R* or a *B* message (columns "Pay *R*" and "Pay *B*", respectively). In addition, the columns "Prob *R*" list the actual relative frequencies of *R* messages conditional on the signal being *R* or *B*.

$s \in \{R,B\}$, i.e. the difference between the values in the first two columns in the panels of Table 6. Further, we use $I_{s=R}$ to denote the indicator checking if the signal is *R* (evaluating to 1 in this case and to 0 otherwise) and allow for logistic errors, implying

$$\Pr(m = R) = \frac{1}{1+\exp\{\lambda \cdot dEP(s) - \kappa \cdot (I_{s=R} - 0.5)\}}. \tag{1}$$

We normalize the indicator by subtracting 0.5. In addition, the differences in expected payoffs of *R* and *B* messages, $dEP(s)$, are estimated from the data and correspondingly subject to measurement error. To account for this, we estimate the standard errors of $dEP(s)$ for each signal $s \in \{R,B\}$ in each treatment and then use the obtained standard errors in an MCMC approach following Hadfield (2010) to correct for measurement error. We account for the panel structure of the data by including random effects at the subject level. The results are provided in Table 7. They are fairly clear and confirm the above impression: the own signal is highly significant in all cases, but the expected payoffs are largely irrelevant. Their coefficients are significant in three out of four treatments, but they are of negligible economic significance in the majority treatments[5] and equally in the unanimity treatments. Hence, we conclude that messages are *non-strategic*, being mostly independent of expected

[5]Note that in the majority treatments, the differences in expected payoffs (the first two columns in the panels of Table 6) are generally less than 2, and hence if weighed with factors less than 1 they are dominated by the signal which has weights at or above 5.

payoffs while correlating strongly with signals

Table 7: Expected payoffs of messages contingent on the private signal

	Majority		Unanimity	
	40-10	35-15	40-10	35-15
Estimates				
Exp. payoff	0.557^{**} (0.161)	0.771 (0.472)	-0.531^{***} (0.092)	-0.234^{**} (0.059)
Own signal	7.295^{***} (0.505)	5.061^{***} (0.493)	5.891^{***} (0.567)	5.987^{***} (0.372)
Additional information				
Number observations	2400	2250	2250	2400
DIC	1442.9	1367.5	957.0	1086.9
Measurment error correction	✓	✓	✓	✓
Random effects	✓	✓	✓	✓

Note: One asterisk indicates p-values $p < 0.05$, two asterisks indicate $p < 0.01$ to approximate the Bonferroni correction (there are four simultaneous tests per hypothesis), and three asterisks indicate $p < 0.001$.

In light of our ex-ante predictions (Table 15), possible explanations are lying aversion and overreaction (besides potentially altruism, which we discuss below). Lying aversion asserts a preference for sending truthful messages, and this preference may outweigh differences in expected payoffs. Overreaction, i.e. overshooting when inferring the state of Nature from signals, strengthens the incentives for sending truthful messages in other ways. On one hand, there is an indirect effect in that co-players overshoot when voting based on the message one sends, and hence sending wrong messages impedes voting efficiency. On the other hand, and this effect can be identified analyzing messages, overreaction dilutes one's payoff expectations.

To clarify, we use $dP(\omega = B)$ to denote the difference of the expected payoffs from messaging R or B given the (unknown) state ω,

$$dP(\omega) = P(m = B|\omega) - P(m = R|\omega), \tag{2}$$

and thus can define the expected difference in payoffs from messaging R and B, given one's

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: This tables provides the conditional payoffs observed in the experiment, conditional on the signal being R or B and the subject sending either an R or a B message (columns "Pay R" and "Pay B", respectively). In addition, the columns "Prob R" list the actual relative frequencies of R messages conditional on the signal being R or B.

belief $\Pr(\omega = R|s)$ about the unknown state of Nature ω given signal s, as

$$dEP(s) = \Pr(\omega = R|s) \cdot dP(\omega = R) + \Pr(\omega = B|s) \cdot dP(\omega = B). \tag{3}$$

Overreaction implies the belief $\Pr(\omega|s)$ to be biased toward one's private signal s, which in turn biases the payoff expectation in favor of messaging truthfully. To illustrate, Table 8 shows the empirically expected payoffs (in the experiment) of messaging R or B contingent on the unknown state of Nature, i.e. the items denoted $P(m = B|\omega)$ and $P(m = R|\omega)$ above. In all treatments, sending the message equating with the unknown state is vastly more profitable than sending the wrong message. Hence, if subjects overestimate the probability that the true state equates with their signal, then they have a strong preference for messaging truthfully.

To verify this point econometrically, let us finally estimate the degree of overreaction, i.e. the beliefs about the true state of Nature, that would rationalize the choices given rational expectations about contingent payoffs $dP(\omega)$. To this end, we capture the mapping from signal s to belief $\Pr(\omega = R|s)$ with a simple logistic function

$$\Pr(\omega = R|s) = \frac{1}{1 + \exp\{\alpha_m \cdot (I_{s=B} - 0.5)\}} = 1 - \Pr(\omega = B|s), \tag{4}$$

which allows us to estimate the weight on the own signal (α_m) alongside precision λ in communication, see Eq. (1), simply by maximum likelihood over observed messages, us-

ing the system of structural equations (1)–(4). Note that Eq. (4) captures potential violations of Bayesian updating in a fairly general manner. Perfect Bayesians have $\alpha_m = 0.79$, which implies the Bayesian posteriors $\Pr(\omega = R|s = R) = 0.6$ and $\Pr(\omega = B|s = R) = 0.4$. Overreaction obtains if $\alpha_m > 0.79$, i.e. subjects overshoot in response to their signal, and conservatism in Bayesian updating obtains if $\alpha_m < 0.79$.

Table 9 summarizes the results. We compare the empirically observed voting probabilities $\hat{\Pr}(m = R|s)$ to the voting probabilities predicted without overreaction ("expected payoffs") and with overreaction. For clarity, we also show the implied beliefs $\Pr(\omega = R|s)$ about the true state ω contingent on the own signal s and the actual parameter estimates (λ, α_m). For brevity, we pool majority and unanimity treatments in Table 9.

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

The weight α_m is consistently around 5 across treatments, vastly in excess of the Bayesian $\alpha_m = 0.79$. Implicitly, subjects' messages are rationalized (consistently across conditions) by the belief the unknown state ω equates with the own signal with probabilities slightly above 0.90. In relation to the Bayesian posterior of 0.60, this indicates possible overreaction to a large degree, and consequently subjects vastly overestimate the payoffs from messaging truthfully. The implied probabilities of messaging *R*, listed in the right-most column of Table 9, closely match the observed relative frequencies across conditions (listed in the left-most column). In absolute terms, the differences between predictions and observations are 3, 9, 1, and 3 percentage points. In contrast, Bayesian updaters would not generally message truthfully, in particular not after *B* signals, and the implied predictions for "expected payoffs" are far off the observed truthfulness after *B* signals.

We reiterate that lying aversion is similarly adequate in rationalizing truthfulness of messages. This point is trivially true in all conditions, it does not even require the state-contingent payoffs in Table 8 to align with truthfulness. That is, both overreaction and lying aversion match the evidence well, and looking at messages alone, differences between these concepts are far from being significant. Analyzing messages does not suffice to disentangle these explanations, but analyzing voting behavior in addition will allow us to progress further, as overreaction also imposes restrictions on voting.

Result 2. *Messages are non-strategic, being mostly independent of expected payoffs while correlating with signals. Lying aversion and overreaction are both compatible with the observed messages.*

5 Do subjects vote rationally?

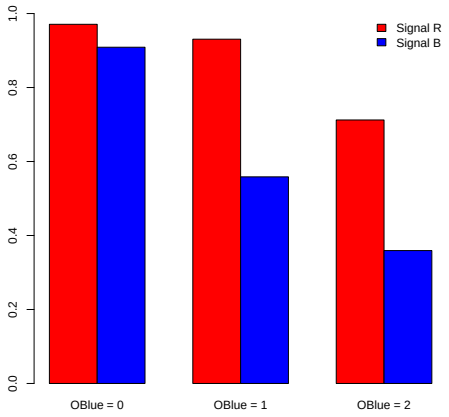
Now, let us turn toward the decisions in the voting stage. Figure 1 provides an overview across information sets. Voting strategies are depicted as functions of the own signal (R or B) and the number of co-players' messages that are B (0, 1 or 2). For both majority and unanimity treatments, subjects vote R with high probability (around 0.9) if the co-players' messages are both R, with relatively low probability if the co-players' messages are both B, and with intermediate probability otherwise. Across information sets, the relative frequency of voting R ranges from 0.1 to 0.95, responding in the intuitive manner to the information available to subjects.

In the following, we analyze to which degree the responsiveness is statistically significant. The pieces of information available to subjects are their own signal s, their own message m_1, and their co-players' messages m_2, m_3. Based on those, a subject's belief about the state of the world can be expressed using the logistic function

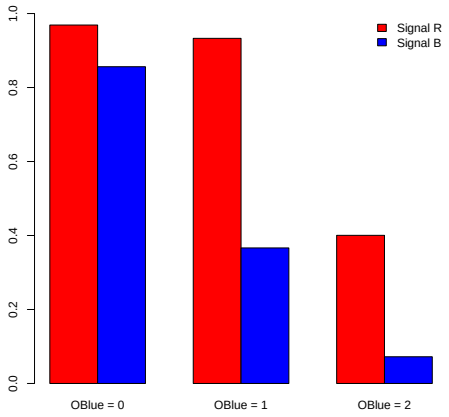
$$\Pr(\omega = R) = \frac{1}{1 + \exp\{\alpha_1 \cdot (I_{s=B} - 0.5) + \alpha_2 \cdot (I_{m_2=B} + I_{m_3=B} - 1)\}}. \tag{5}$$

Figure 1: Relative frequency of voting R as functions of own signal and the number "OBlue" of blue messages sent by co-players

(a) Majority treatments



(b) Unanimity treatments



By regressing the unknown state of the world on the available information, we estimate the weights a subject with rational expectations would put on each of these pieces of information when predicting the state. Similarly, we can represent the probability $\Pr(X = R|v)$ that the voting outcome X is R based on the available pieces of information (including the own vote v). Our preferred specification is as follows.

$$\Pr(X = R|v) = \frac{1}{1+\exp\{\beta_1 \cdot I_{v=B} + \beta_{2|0} I_{m_2=m_3=R} + \beta_{2|1} I_{m_2\neq m_3, m_1=R} + \beta_{2|2} I_{m_2\neq m_3, m_1=B} + \beta_{2|3} I_{m_2=m_3=B}\}}, \tag{6}$$

and the complemetary probability that B results is $\Pr(X = B|v) = 1 - \Pr(X = R|v)$. Thus, the voting outcome is expressed as a logistic function of the own vote v (weight β_1) and of the co-player's messages in relation to the own message. In our preferred specification, we allow for four cases: the co-players' messages are either both R, both B, or mixed, and in the latter case, we distinguish two cases based on the own signal (since this signal is visible to the co-players). In general, six cases are possible, as discrimination based on the own signal would be possible also in the two cases with unanimous messages of co-players, but the adequacy of the representation does not improve significantly when we do so. In this sense the extension is not needed to express rational expectations about the voting outcome

and can be skipped to maintain parsimony.

Now, with these expressions of rational expectations about state and voting outcome in hand, it is straightforward to compute the expected payoffs given rational expectations. The probability that the voting outcome equates with the true state of Nature is

$$\Pr(X = \omega | v) = \Pr(\omega = R) \cdot \Pr(X = R \,|\, v) + \Pr(\omega = B) \cdot \Pr(X = B \,|\, v). \tag{7}$$

Thus, using P_c and K to denote the payoff from the voting outcome being correct and the bonus from voting R, respectively, the expected payoffs of voting $v \in \{R, B\}$ are expressed straighforwardly. In majority treatments, we obtain

$$EP(v = R) = P_c \cdot \Pr(X = \omega | v = R) + K \qquad EP(v = B) = P_c \cdot \Pr(X = \omega | v = B), \tag{8}$$

and in unanimity treatments, we obtain for both $v \in \{R, B\}$,

$$EP(v) = P_c \cdot \Pr(X = \omega | v) + K \cdot \Pr(X = A | v). \tag{9}$$

The belief weights in Eqs. (5) and (6) are simply estimated by maximum likelihood and with these estimates, we obtain estimates of the expected payoffs using Eqs. (7)–(9). We bootstrap the distribution of the belief weights obtained by regression, stratifying at the subject level to account for the panel structure of the data, and by recomputing the expected payoffs for each bootstrap sample, we obtain estimates for the standard errors of expected payoffs across information sets. Table 10 presents the results, and to put the estimates into context, it also includes the observed relative frequencies of voting R.

Two observations are striking. First, in all information sets of both majority treatments, subjects would maximize expected payoffs by voting R. The differences in expected payoffs are always close to the payoff from voting expressively, 10 or 15 points, suggesting that subjects actually have little impact on the voting outcome (i.e. on the probability that the voting outcome equates with the state of the world). In line with this observation, a majority

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: The tables show, for each information set in the voting stage, the expected payoff from voting R, denoted $EP(R)$, the expected payoff from voting B, denoted $EP(B)$, and the relative frequency of R votes, denoted $\hat{\Pr}(R)$. These relative frequencies are set in bold-face type if the observation contradicts payoff maximization, and specifically if subjects predominantly vote B although expected payoffs predict R to be more profitable.

of subjects votes B only in three out of the 24 information sets. Second, the differences in expected payoffs are much smaller in unanimity treatments, in particular if signals and messages suggest that R should result—as B can be vetoed. If at least one subject on each committee sticks to voting R, then R results, and both other subjects are indifferent between R and B. In contrast, if B is the likely state of the world, then the decision of voting R or B is payoff relevant and subjects indeed are best off voting B. Again, however, subjects tend to vote B too often, i.e. even in some of the information sets where voting A is optimal. They do so when signals and messages suggest they should, but this tendency is incompatible with rational expectations about expected payoffs.

Econometrically, we verify the relevance of expected payoffs in voting by regressing votes on signals and messages, and by additionally including the difference in expected payoffs as reported in Table 10. Under rational expectations, the expected payoffs should be of significant relevance, as they contain all relevant information, while signals and messages should be insignificant. As above, the estimated expected payoffs are subject to measurement error, which we account for using the MCMC approach of Hadfield (2010) using the bootstrapped standard errors reported in Table 10, and we include random effects at the sub-

ject level to account for the panel structure of the data. The results are reported in Table 11.

Table 11: Relevance of expected payffs, signal and co-players' messages in voting

	Majority		Unanimity	
	40-10	35-15	40-10	35-15
Estimates				
Exp. Payoff	-0.083 (0.055)	-0.023 (0.048)	0.337^{***} (0.007)	0.152^{***} (0.038)
Own signal	1.804^{***} (0.327)	2.452^{***} (0.179)	4.752^{***} (0.019)	3.564^{***} (0.112)
Opp's messages	1.403^{***} (0.219)	1.607^{***} (0.167)	3.585^{***} (0.016)	2.771^{***} (0.116)
Constant	-3.122^{***} (0.821)	-4.732^{***} (0.879)	-7.268^{***} (0.217)	-6.2^{***} (0.273)
Additional information				
Number observations	2400	2250	2250	2400
DIC	1985.6	1608.0	1347.0	1576.4
Measurment error correction	✓	✓	✓	✓
Random effects	✓	✓	✓	✓

Note: Logit regression with random effects at subjects level, across all rounds per session; robustness checks for each half session could go into supplement/appendix; exp. payoff is $EP(v = R) - EP(v = B)$, DIC is the deviance information criterion

To begin with, the own signal and the co-players' messages are of similar relevance. The own signal gets higher weight than the co-players' messages individually, but jointly their messages have higher weight than one's signal. This corresponds with the general level of truthfulness, which is not quite 100% and implies that the opponents' messages need to be discounted, indeed. In contrast, expected payoffs are entirely insignificant in the majority treatments, and significant but economically of minor relevance in the unanimity treatments. The difference in expected payoffs has the highest weight in the 40-10 unanimity treatment (weight 0.337), and the difference itself is generally less than 10 points across information sets in this treatment. The maximal impact on choice propensities by expected payoffs is therefore less than 3.37, mostly substantially less, and as such it is still much less relevant in choice than any co-player message, let alone the own signal. This obtains, even though the rationally choice relevant information that is contained in messages and signals already has been accounted for in the computation of expected payoffs.

Result 3. *Subjects message truthfully, slightly discount the co-players' messages, vote re-*

sponsively, but do not consistently vote for the option that would maximize their individual payoffs. They tend to vote for the option supported by the majority of signals and messages.

This over-weighting of the information contained in messages and signals again suggests that subjects exhibit an overreaction, indeed it is predicted by the above findings from analyzing messages. An alternative explanation is, however, that subjects do not hold rational expectations about their co-players' voting strategies—as proposed by level-k (and cursedness). For, intuitively, if one overestimates the noise in the co-players' voting, then one also overestimates the probability that one's vote will be pivotal, and consequently, one overestimates the expected gain from voting in line with the majority of signals and messages.

6 Why do subjects deviate from rational behavior?

The previous sections have shown that the exceeding efficiency of information aggregation in committees relates to two basic observations: Committee members communicate more truthfully than predicted by expected payoffs, and they vote more sincerely than predicted by expected payoffs, over-weighting signals and messages in relation to the expected payoffs actually implied by signals and messages. Lying aversion can explain truthful messages but not sincere voting. Limited depth of reasoning can partially explain sincere voting, as it biases payoff expectations, but it does not explain the high weights subjects put on their co-players' messages—those would be perceived as uninformative by players with limited depth of reasoning. Overreaction is in principle compatible with both observations, but the above analysis left open if the degree of overreaction is consistent in messaging and voting, and sufficient to jointly explain both aspects. The purpose of the present section is to econometrically disentangle and test these potential explanations.

To this end, we merge the models of voting and messaging introduced above. On one hand, using the expected payoffs of voting $v \in \{A,B\}$ as defined above, $EP(v = R)$ and $EP(v = B)$ following Eqs. (6)–(9), and allowing for logistic errors, the subject will vote R

with probability

$$\Pr(v = R) = \frac{1}{1 + \exp\{-\lambda_1 \cdot (EP(v = R) - EP(v = B)) - \lambda_2 \cdot K\}} \tag{10}$$

Initially, we allow for $\lambda_1 \neq \lambda_2$, i.e. for subjects to exhibit efficiency concerns ($\lambda_1 > \lambda_2$) or to otherwise deviate from equally weighing the two payoff dimensions, but this will prove to be an unnecessary degree of caution. Following the above result, that messages following R signals tend to be more likely to be truthful than messages following B signals, we extend the model introduced above by allowing for such asymmetry also in the own belief updating.

$$\Pr(\omega = A) = \frac{1}{1 + \exp\{\alpha_m(I_{s=B} - 0.5) + \alpha_1 \cdot I_{m_2=R \vee m_3=R} + \alpha_2 \cdot I_{m_2=B \vee m_3=B}\}} = 1 - \Pr(\omega = B) \tag{11}$$

By Eqs. (6)–(9), the weights $\alpha_m, \alpha_1, \alpha_2$ as well as $\beta_1, \beta_{2|\cdot}$ are free parameters characterizing Bayesian updating ($\alpha_m, \alpha_1, \alpha_2$) as well as beliefs about voting strategies ($\beta_1, \beta_{2|\cdot}$).

On the other hand, using expected payoffs from communicating given the own signal, $dEP(s)$ as defined in Eqs. (2)–(4) with free parameters $(\alpha_m, dEP(R), dEP(B))$, the subject will send message R with probability

$$\Pr(m = R|s) = \frac{1}{1 + \exp\{\lambda_3 \cdot dEP(s) - \lambda_4 I_{s=A} + \lambda_4 I_{s=B}\}}. \tag{12}$$

Besides allowing for stochastic errors, this formulation allows us to test for the significance of lying aversion ($\lambda_4 > 0$) and incorrect Bayesian updating ($\alpha_m \neq 0.79$). In relation to above, we make the identifying assumption that messaging agents hold rational expectations about the expected payoffs from R or B messages conditional on the state of the world, i.e. $dP(\omega)$ are anticipated correctly. As indicated, agents may deviate from truthful communication because of wrong Bayesian updating or lying aversion.

We start the analysis with a general model of behavior containing all parameters α_1, α_2, β_1, $\beta_{2|\cdot}$ and λ_1–λ_4 as degrees of freedom. In the analysis, we will successively restrict this

general model of behavior, in order to identify the biases actually present in behavior. That is, we determine the weights α_1, α_2 as well as $\beta_1, \beta_{2|\cdot}$ actually used by subjects, test whether and how subjects deviate from rational expectations and Bayesian updating, and restrict the model when appropriate. In this way, starting at the general model and successively identifying the behaviorally relevant biases, we can analyze behavior without making inadequate assumptions at any step in the analysis.

All parameters are estimated by maximum likelihood, standard errors are Huber-Sandwich estimates clustered at the subject level, and the hypothesis tests discussed next are implemented as likelihood ratio tests evaluating the general model against specifically restricted models, bootstrapping the test statistic by replacement again at the subject level (to control for the panel character of the data). Table 12 presents the parameter estimates discussed in the following. In addition, Table 12 presents the results of all likelihood ratio tests investigating significance of specific biases, to be discussed, and the empirical estimates $\hat{\alpha}$ and $\hat{\beta}$ that represent the respective values if subjects held rational expectations and obeyed Bayesian updating.

We start by estimating the model parameters for each treatment in isolation, either for all periods, or separately for the first half of periods ($1 \dots 25$) and the second half of periods ($26 \dots 50$). In likelihood ratio tests, we evaluate the differences between the two unanimity treatments, on the one hand, and the two majority treatments, on the other hand. They prove to be far from significant, the p-values being above $p = 0.8$ at all levels of experience and also pooled across all rounds, which shows that we can pool the two majority treatments and two unanimity treatments, respectively; naturally, we control for the treatment differences in expressive payoffs K. Based on this, we test for learning effects. If each treatment is considered in isolation, behavior does not differ significantly between first and second halves of the sessions, the p-values being around $p = 0.1$ in all four treatments, but when we pool majority treatments and unanimity treatments, respectively, the experience effects exceed the conventional level of significance, with the p-values being $p = 0.027$ and $p = 0.035$ in majority and unanimity treatments (respectively). That is, with experience, behavior adapts,

Table 12: Analysis of the motives underlying communication and voting

Null Hypothesis	Alternative	Majority decision				Unanimous decision			
		First half		Second half		First half		Second half	
		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) $H_E : H_D \wedge \beta_{2\|0...3} = \hat{\beta}_{2\|0...3}$	H_D		32.69^{***} (0)		30.44^{***} (0)		340.34^{***} (0)		301.08^{***} (0)
(F) $H_F : H_D \wedge \beta_{2\|0...3} \propto \hat{\beta}_{2\|0...3}$	H_D		15.09^{**} (0.019)		9.35^{*} (0.088)		31.66^{**} (0.036)		8.94 (0.3)
(G) $H_G : H_D \wedge \beta_{2\|0...3} = 0$	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 reports the parameters estimated by maximum likelihood, the standard errors reported in parentheses are Huber-Sandwich estimates clustered at the subject level. For each hypothesis test, we present the logarithm of the likelihood ratio of restricted model over general model and the bootstrapped p-value in parentheses, and asterisks to indicate the level of significance: one asterisk indicates significance at the 0.1 level, and two asterisks indicate significance at 0.01. H_{Base} represents the hypothesis that all parameters are free, and $\hat{\alpha}, \hat{\beta}$ refer to the empirical estimates. The notation $\alpha_{1,2}$ refers to the vector (α_1, α_2), and the symbol $\propto$ indicates proportionality of one vector to another.

presumably behavioral biases diminish in relevance. For this reason, we do not pool the data across all rounds, but focus on the results for each half session in isolation.

The question to be answered now is which of the biases actually are indeed significant. First, we verify if subjects equally weigh expected payoffs in voting and message stage ($\lambda_1 = \lambda_3$), and simultaneously if subjects over- or underweigh expressive payoffs ($\lambda_1 = \lambda_2$). Efficiency concerns have been observed by Charness and Rabin (2002), Engelmann and Strobel (2004) and many subsequent studies, and applied to voting games, this may lead subjects to overweigh the payoff from getting the state right. Overweighing expressive payoffs can also be interpreted as a tribute to bounded rationality, as securing the bonus payment is a much less complex problem to solve than inferring the state of the world.

Question 1. *Do subjects weigh expressive payoffs proportionally ($\lambda_1 = \lambda_2 = \lambda_3$)? – Yes*

We test for efficiency concerns in a two-sided test of the null hypothesis $\lambda_1 = \lambda_2 = \lambda_3$, against the alternative that the unrestricted model H_{Base}, allowing for arbitrary $(\lambda_1, \lambda_2, \lambda_3)$, better represents behavior. The results, displayed in row (A) in Table 12, show that the null is maintained with high p-values, being above 0.2 in all four data subsets. Thus, there is no indication that subjects actually deviate from weighing the three payoff dimensions asymmetrically. Given this, let us turn to analyzing belief updating about the state of Nature.

Question 2. *Do subjects hold Bayesian beliefs about the true state of Nature? – No*

Subjects have Bayesian beliefs about the true state of Nature if they correctly weigh their own signal ($\alpha_m = \hat{\alpha}_m$) and their opponents' messages ($\alpha_{1,2} = \hat{\alpha}_{1,2}$), which in turn requires rational expectations about the opponents' strategies mapping signal to messages. If subjects hold rational beliefs, or act behaviorally equivalently, then imposing these restrictions on parameters will not induce a significant drop in the log-likelihood. Row (B) in Table 12 reports on the results of this test, showing that subjects significantly deviate from rational beliefs in all four (sub-) sets of the data, with p-values ranging from .001 (inexperienced subjects under unanimity) to .042 (experienced subjects under unanimity) This raises the

question how subjects fail to form rational beliefs. The most prominent such conjectures, limited depth of reasoning or cursedness (Eyster and Rabin, 2005), is tested next.

Question 3. *Do subjects have cursed beliefs about the true state of Nature? – No*

Cursedness and similarly level-1, asymmetric logit equilibrium, and noisy introspection posit that players believe their opponents' messages are less informative than they actually are. Subjects put therefore a relatively higher weight on their own signal (α_m) than on the opponents' messages ($\alpha_{1,2}$) judged in relation to the their truthfulness. Correspondingly, the ratios $\alpha_m/\alpha_{1,2}$ are greater than $\hat{\alpha}_m/\hat{\alpha}_{1,2}$, which we again test in a two-sided test, thus allowing for deviations also in the alternative direction as control. The results are reported in row (C) in Table 12, and the likelihood ratios are insignificant across all treatments and levels of experience. For example, in the first session halves on majority treatments, the rational weights are $(\hat{\alpha}_m, \hat{\alpha}_1, \hat{\alpha}_2) = (.771, .574, .377)$, and the subjects' weights are $(\alpha_m, \alpha_1, \alpha_2) = (22.6, 12.2, 12.0)$. These weights differ significantly, as observed above, but their ratios do not differ notably (let alone significantly) in the sense that these vectors are proportional to one another. That is, subjects do not exhibit cursed beliefs, and they do not ignore the rationality of others in other respects. They simply overshoot (rather drastically) when forming beliefs about the true state of Nature based on their information.

Next, we turn to analyzing expectations about the voting outcome.

Question 4. *Do subjects hold rational expectations about the impact of their own vote? – Yes*

Line (D) reports the test of our next hypothesis that subjects hold rational expectations of the relevance of their own vote for the ultimate voting outcome. The null hypothesis $\beta_1 = \hat{\beta}_1$ is never rejected significantly, and in the most extreme case, experienced subjects in unanimity, the *p*-value of this null hypothesis is .106. This cannot be called weakly significant, specifically not in light of the multiple testing problem we are analyzing. Next, we ask the complementary question whether subjects hold rational expectations about their opponents' voting behavior conditional on their messages.

Question 5. *Do subjects hold rational expectations about their opponents' voting? – No*

Rational expectations about the (conditional) voting outcome requires correct inference about the other votes from the list of messages, which requires Bayesian updating about the opponents' signals, given their messages, and rational expectations about their voting strategies given signals and messages, plus adequate weighting of the known own vote in relation to the probabilistic beliefs about the opponents' votes. Using the notation introduced above, subjects hold rational expectations if $\beta_{1,2|0...3} = \hat{\beta}_{1,2|0...3}$, which we again verify in (two-sided) likelihood ratio tests. The results are reported in row (E) in Table 12, showing that subjects singificantly violate rational expectations in all four cases. The deviation from rational expectations is probably not very surprising and leads to the more subtle question of how subjects actually deviate from rational expectations. We test three possible conjectures.

Question 6. *Do subjects over-/undershoot forming beliefs about the voting outcome? – Yes, when they are experienced*

If subjects merely over- or undershoot forming beliefs about the voting outcome, based on their own vote and the message profile, then $\beta_{2|0...3}$ is proportional to $\hat{\beta}_{2|0...3}$, i.e. a linear transformation thereof. We test for proportionality in likelihood ratio tests again, row (F) in Table 12 reports the results. The null hypothesis of proportional weights, i.e. unbiased beliefs, is rejected when subjects are inexperienced (where the beliefs deviate widely from rational expectations), but not anymore when subjects are experienced.

This suggests that subjects may hold rational expectations about the voting strategies and simply over- or undershoot when predicting the voting outcome based on their opponents' messages, perhaps because of having biased beliefs as hypothesized by cursed equilibrium and level-k theories. Subjects with cursed beliefs have rational expectations about their opponents' voting strategies but underestimate the predictability of their opponents' votes conditional on their messages. In this case, $\beta_{1,2|0...4}$ will not be proportional to $\hat{\beta}_{1,2|0...3}$, while $\beta_{2|0...3}$ is proportional to $\hat{\beta}_{2|0...3}$ and only scaled down in relation to β_1. This corresponds with the observation that we just made. As a control, let us next test the more extreme hypothesis that the beliefs about the opponents' voting strategies actually correspond with

level-1 beliefs (that opponents randomize uniformly during voting): $\beta_{2|0...3} = 0$.

Question 7. *Do subjects have level-1 beliefs when voting? – Yes, when they are experienced*

If subjects have limited depth of reasoning in the sense of level-k, and in particular if they hold level-1 beliefs, then they expect opponents to randomize uniformly when sending messages or voting. Formally, level-1 therefore predicts $\beta_{2|\cdot} = 0$, and as shown in row (G) in Table 12, level-1 behavior is rejected when subjects are inexperienced, but not anymore when subjects are experienced – essentially at the same levels as we observed for cursed beliefs before. That is, the apparent learning that we observed is that subjects stop holding wrong beliefs about their opponents' votes and transition to very noisy beliefs. The result is that subjects vote as if believing that their vote will be pivotal with .5 probability (against uniformly randomizing opponents), which greatly inflates their pereption of the pivotality of their vote. Notably, however, subjects do not hold level-1 beliefs about their opponents' messages, i.e. they are not textbook level-1 players. For this reason, we shall refer to this behavioral bias as pivotality illusion in this paper.

Question 8. *Is lying aversion a significant behavioral factor? – No*

The results of our test of lying aversion ($\lambda_4 = 0$) are reported in row (H) of Table 12. Row (H) shows that lying aversion is not generally a significant factor, but it is weakly significant for experienced subjects in the majority treatments. In this case, the subjects' messages are entirely independently of the expected payoffs empirically associated with the messages, as indicated by the estimated $\lambda_3 = 0$, and as result, voting strategies can be explained only referring to lying aversion. In the grand scheme of things, this single detection of a weak significance ($p = 0.058$) in this multiple-testing analysis with four tests is not robust to a Bonferroni correction, suggesting that lying aversion is actually not a behavioral factor in our experiment.

Importantly, this obtains once we allow for imperfect Bayesian updating. When we were to assume perfect Bayesian updating, lying aversion would be significant—but perfect Bayesian updating was rejected highly significantly even allowing for the other biases, lim-

ited depth of reasoning (pivotality illusion) and lying aversion. In contrast, lying aversion is not significant when allowing for the other biases. Hence, imperfect Bayesian updating is an inevitable part of the explanation, and lying aversion is not.

Result 4. *In voting games with communication, subjects mainly violate Bayesian updating: they overshoot when forming beliefs about the true state of Nature ("overreaction") and undershoot when predicting the voting outcome ("conservatism").*

1. *Overreaction is a highly significant behavioral bias even when we allow for lying aversion and limited depth of reasoning.*

2. *Once we allow for overreaction, lying aversion is not significant.*

3. *Limited depth of reasoning helps explain voting behavior, but it is not compatible with the (large) extent to which subjects infer information from messages, suggesting that this behavioral bias represents a "pivotality illusion".*

Result 4 provides a qualitative assessment of the behavioral factors, which we complement by a quantitative assessment next. That is, having seen that overreaction seems to be a substantial factor, how much can we actually explain by overreaction, and is it important to account for limited depth of reasoing (pivotality illusion) in addition? In turn, lying aversion does not seem to be a substantial factor, in the sense that overreaction explains all that lying aversion explains and more, but how much more is it actually? In order to answer these questions, we look at the models allowing for limited depth of reasoning ($\beta_{2|0\ldots3} \propto \hat{\beta}_{2|0\ldots3}$), lying aversion ($\lambda_4 \neq 0$), and overreaction ($\alpha_{m,1,2} \propto \hat{\alpha}_{m,1,2}$)—either in isolation, or in conjunction, or not all. The model allowing for all biases in conjunction provides an upper bound ($LL_{\max}$) of the log-likelihood that we can reach given our understanding of behavioral biases. The model allowing for none of the biases merely allows for logit response to rational expectations, corresponding to the logit QRE of McKelvey and Palfrey (1998). The models allowing for any one of the behavioral biases in isolation informs us how big a factor this particular bias is, in relation to the full model and the unbiased QRE. Based on this, we quantify the explanatory power of each model using the pseudo-R^2 (Nagelkerke, 1991),

$R^2 = \frac{LL - LL_{\min}}{LL_{\max} - LL_{\min}}$, where $LL_{\min}$ is the absolute lower bound attained by predicting uniform randomization for all decisions, which allows us to also put QRE into perspective.

Table 13: 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	0.9	1
+ lim dep reasoning	-5102.9	0.49	1	0	0	-4715.25	0.57	1	0	0.19

Note: γ_1 denotes the degree of overreaction (formally $\alpha_1/\hat{\alpha}_1$), γ_2 denotes the degree of of lying aversion (formally λ_4/λ_1), and γ_3 denotes the depth of reasoning (formally $\beta_{2|0}/\hat{\beta}_{2|0}$).

Table 13 reports the results, distinguishing inexperienced and experienced subjects in our experiment, i.e. first and second halves of sessions. QRE on its own explains 35% and 47% of observed noise for inexperienced and experienced subjects (respectively). Limited depth of reasoning adds 10–15 percentage points in either case, suggesting that it is of minor economic relevance despite the statistical significance observed above. Lying aversion explains about 20 percentage points more than limited depth of reasoning in either case, putting the explanatory power of limited depth of reasoning further into perspective. Overreaction, however, pushes QRE's explanatory power to 87% robustly for both inexperienced and experienced subjects. This show that overreaction, a simple model allowing for the vector $\alpha_{m,1,2}$ to be proportional to its Bayesian counterparts $\hat{\alpha}_{m,1,2}$, captures a large fraction of observed noise on its own (40–50 percentage points), raising QRE's explanatory power to 87% for both inexperienced and experienced subjects.

Result 5. *To a large extent (87%), behavior is well-approximated by a tractable, one-parametric generalization of quantal response equilibrium allowing for overreaction.*

In turn, QRE with overreaction leaves 13 percentage points to be explained when applied in addition to logistic errors (QRE). As seen above, the residual 13 percentage points of observed noise amount largely to limited depth of reasoning, but in the grand scheme of things, it of comparably minor economic relevance.

7 Conclusion

We report the results of an experiment on committee decision making where committee members communicate via formalized cheap talk, vote for a committee decision under either majority or unanimity, and may collect either high or low expressive payoffs when voting for one of the options. The novelty of this 2×2 design is that it separates the leading theories about committee behavior (lying aversion and limited depth of reasoning) from each other and from a third widely documented bias known as overreaction in response to news. Our central observations are that subjects communicate more truthfully than justified by expected payoffs, they correctly anticipate the imperfect truthfulness of their co-players' messages, and they overweigh private signals and public messages in the voting stage – which contradicts limited depth of reasoning as well as lying aversion, but has been predicted by overreaction. As simple model generalizing logit equilibrium by a single parameter quantifying this overreaction explains 87% of observed behavior in our experiment.

Qualitatively, this extends directly to the array of experiments with seemingly diverging observations discussed in the introduction. Most notably, Goeree and Yariv (2011) analyze voting in committees with pre-play communication and similarly find overcommunication. In their setting, overcommunication is implied by overreaction to news (signals) for the same reason as it is in our framework, which differs only by additionally allowing for expressive payoffs in voting. If an agent overestimates the information contained in their signal, then she will overestimate the importance of communicating it correctly. Le Quement and Marcin (2020) extend Goeree and Yariv's framework by allowing for publicly known diverse preference and similarly to us, they find that a large majority of subjects communicates truthfully and votes with the the majority of communicated messages. While our framework differs substantially—allowing for both expressive payoffs and unanimity voting—our model allowing for overreaction intuitively applies equally in their framework, by predicting the same set of behaviors and thus providing a microfoundation also for the findings of Le Quement and Marcin (2020). Most recently, Ginzburg et al. (2022) analyze committee voting about donations to charity, focusing on comparative statics of behavior with respect to the

pivotality of individual committee members and seeking to examine if subjects implicitly hold expressive payoffs when voting about charitable donations. Their results are in the affirmative, thereby providing a foundation for explicitly inducing expressive payoffs as in our experiment, which in turn was the key to our identification of overreaction.

To our knowledge imperfect Bayesian updating has not been discussed in the context of committee behavior, but outside committee behavior, a large literature has established behavioral overreaction of subjects as new information arrives (Kahneman and Tversky, 1973; Tversky and Kahneman, 1982; Bar-Hillel, 1980; Bordalo et al., 2020; Afrouzi et al., 2023) and its relevance in many applied contexts such as auctions and signaling games (Eyster and Rabin, 2005). While Eyster and Rabin's concept cursed equilibrium implies that subjects tend to underestimate the information contained in the actions of others, which is not compatible with our observations in committee behavior, the underlying idea of imperfect Bayesian updating works similarly. The existing literature contains several promising models of non-Bayesian updating (Epstein, 2006; Epstein et al., 2008; Ortoleva, 2012; Massari, 2021) and more recently asymmetric updating (De Filippis et al., 2022), all of which relax the axiomatic foundation of Bayesian updating (Alós-Ferrer and Mihm, 2023) in one way or the other and thus are potential models for future analyses of committee behavior. Our experiment was not designed to discriminate between these models, being designed to discriminate between lying aversion, limited depth of reasoning and imperfect Bayesian updating. Based on our results and the models developed in this literature, however, allowing for imperfect Bayesian updating of committee members has great potential for modeling committee behavior, towards a better understanding of the optimal design of voting rules and communication structures in committees.

References

Abeler, J., Nosenzo, D., and Raymond, C. (2019). Preferences for truth-telling. *Econometrica*, 87(4):1115–1153.

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.

Alós-Ferrer, C. and Mihm, M. (2023). An axiomatic characterization of bayesian updating. *Journal of Mathematical Economics*, 104:102799.

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

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

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. (2022). Strategic communication in committees with mixed motives. *The RAND Journal of Economics (accepted for publication)*.

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.

Charness, G. and Dufwenberg, M. (2010). Bare promises: An experiment. *Economics Letters*, 107(2):281–283.

Charness, G. and Dufwenberg, M. (2011). Participation. *American Economic Review*, 101(4):1211–1237.

Charness, G. and Rabin, M. (2002). Understanding social preferences with simple tests. *The Quarterly Journal of Economics*, 117(3):817–869.

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

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). Non-bayesian updating in a social learning experiment. *Journal of Economic Theory*, 199:105188.

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

Ellingsen, T., Johannesson, M., Tjøtta, S., and Torsvik, G. (2010). Testing guilt aversion. *Games and Economic Behavior*, 68(1):95–107.

Engelmann, D. and Strobel, M. (2004). Inequality aversion, efficiency, and maximin preferences in simple distribution experiments. *The American Economic Review*, 94(4):857–869.

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

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

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

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

Feddersen, T. and Pesendorfer, W. (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.

Hurkens, S. and Kartik, N. (2009). Would i lie to you? on social preferences and lying aversion. *Experimental Economics*, 12(2):180–192.

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

Kawagoe, T. and Takizawa, H. (2008). Equilibrium refinement vs. level-*k* analysis: An experimental study of cheap-talk games with private information. *Games and Economic Behavior*.

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.

Lundquist, T., Ellingsen, T., Gribbe, E., and Johannesson, M. (2009). The aversion to lying. *Journal of Economic Behavior & Organization*, 70(1):81–92.

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

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

McKelvey, R. D. and Palfrey, T. R. (1998). Quantal response equilibria for extensive form games. *Experimental economics*, 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.

Nagelkerke, N. J. (1991). A note on a general definition of the coefficient of determination. *Biometrika*, 78(3):691–692.

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

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.

Tversky, A. and Kahneman, D. (1982). Evidential impact of base rates. *Judgment under uncertainty: Heuristics and biases*, pages 153–160.

Vanberg, C. (2008). Why do people keep their promises? an experimental test of two explanations. *Econometrica*, 76(6):1467–1480.

A Efficient limiting-logit equilibria across treatments

We compute the sequential equilibria numerically as limiting logit agent quantal response equilibria, following McKelvey and Palfrey (1998). This section presents the sequential equilibria that can be expressed as limiting-logit equilibria (precision $\lambda = 10^6$) and are socially efficient within the set of equilibria. The notation of the equilibria is fairly self-explanatory. On the one hand, the message strategy $\sigma()$ has two components, namely the probabilities to send message R after the two possible signals (R and B, in that order). On the other hand, for each combination of signal s_i and message m_i, the voting strategy $\tau(s_i, m_i,)$ has N components: the probabilities of voting for R given the number of players voting in favor of B, i.e. for the N possible values of $A_i = \#\{j|m_j = b\}$ in increasing order. For clarity, the possible values are $(1,2,\ldots,N)$ if i sent message B, and $(0,1,\ldots,N-1)$ if i sent R.

There are two pieces of additional information: π denotes the expected payoffs of the (symmetric) players implied in equilibrium and c denotes the probability that the decision of the committee coincides with the majority of (private) signals of the committee members.

Expected payoffs in the various information sets

For clarity, we spell out the payoffs as functions of the strategy. The notation is as follows:

- Players $i \in \{1,\ldots,n\}$, state of the world ω, signal profile $s = \{s_1,\ldots,s_n\}$, message profile $m = \{m_1,\ldots,m_n\}$, vote profile $v = \{v_1,\ldots,v_n\}$
- Terminal payoff $p_i(\omega, v, m, s) \equiv p_i(\omega, v, m_i, s_i)$, of player i, as a function of above terms, and under our assumptions it is even independent of (m_i, s_i)
- Message strategy $\Pr(m_i|s_i)$ assigns probabilities to messages $m_i \in \{R, B\}$ conditional on signal s_i
- Voting strategy $\Pr(v_i|m, s_i)$ assigns probabilities to votes $v_i \in \{R, B\}$ conditional on signal s_i and message profile m. In the representation above, payoff equivalent information sets are merged. In particular, all message profiles corresponding to a given own signal m_i and a number of aggregate R-signals ($\#R$) are payoff equivalent. We denote "the set of all message profiles m corresponding to some $m_i, \#R$" as $m \hat{=} \#R, m_i$.

Preliminaries Conditional expectation $E(X|Y=y) = \sum_x x \cdot \frac{P(X=x, Y=y)}{P(Y=y)}$. Here and in the following, a comma (",") indicates "logical and". To further simplify notation, we suppress the distinction between random variables and realizations, as the expressions would otherwise be too messy. For example, the above would be written as $E(x|y) = \sum_{x'} x' \cdot \frac{P(x', y)}{P(y)}$.

Terminal node probabilities The probability of reaching node (ω, s, m, v) is

$$\begin{aligned}\Pr(\omega, v, m, s) &= \Pr(\omega) \cdot \Pr(s|\omega) \cdot \Pr(m|s) \cdot \Pr(v|m, s) \\ &= \Pr(\omega) \cdot \prod_{i \le n} \Pr(s_i|\omega) \cdot \prod_{i \le n} \Pr(m_i|s_i) \cdot \prod_{i \le n} \Pr(v_i|m, s_i),\end{aligned}$$

i.e. a function of game parameters, signaling strategy, and voting strategy.

Expected payoffs in voting stage A player votes in information sets (m, s_i) or Markovian in $(\#A, m_i, s_i)$. Thus, we are looking for $E(p|v_i, \#A, m_i, s_i)$, i.e. the expectation of payoff p conditional on vote v_i in state $(\#A, m_i, s_i)$. First, as an illustration, we define a few simpler terms and successively merge nodes to derive at the expectation in the relevant information sets.

First, the expectation in a terminal node is given (and trivial).

$$E(p|v, m, s, \omega) = p(v, m, s, \omega)$$

Second, merging across (unobserved) states ω yields

$$E(p|v, m, s) = \sum_{p'} p' \cdot \frac{\Pr(p = p', v, m, s)}{\Pr(v, m, s)} = \sum_{\omega} p(v, m, s, \omega) \cdot \frac{\Pr(\omega, v, m, s)}{\Pr(v, m, s)}$$
$$= \frac{\sum_{\omega} p(v, m, s, \omega) \cdot \Pr(\omega, v, m, s)}{\sum_{\omega} \Pr(\omega, v, m, s)}$$

Note that $\Pr(v, m, s)$ can be equivalently defined as

$$\Pr(v, m, s) = \sum_{\omega} \Pr(\omega, v, m, s) \equiv \sum_{\omega} \Pr(v, m, s|\omega) \cdot \Pr(\omega),$$

above, the first expression is used. Now, merging across opponent votes and next across signals (again, "$v \hat{=} v_i$" denotes all vote profiles v compatible with the own vote v_i)

$$E(p|v_i, m, s) = \sum_{\omega} \sum_{v \hat{=} v_i} p(v, m, s, \omega) \cdot \frac{\Pr(\omega, v, m, s)}{\Pr(v_i, m, s)}$$
$$E(p|v_i, m, s_i) = \sum_{\omega} \sum_{v \hat{=} v_i} \sum_{s \hat{=} s_i} p(v, m, s, \omega) \cdot \frac{\Pr(\omega, v, m, s)}{\Pr(v_i, m, s_i)}$$
$$= \frac{\sum_{\omega} \sum_{v \hat{=} v_i} \sum_{s \hat{=} s_i} p(v, m, s, \omega) \cdot \Pr(\omega, v, m, s)}{\sum_{\omega} \sum_{v \hat{=} v_i} \sum_{s \hat{=} s_i} \Pr(\omega, v, m, s)}$$

Finally, merging to reach the Markovian information sets, we get

$$E(p|v_i, \#A, m_i, s_i) = \sum_{\omega} \sum_{v \hat{=} v_i} \sum_{m \hat{=} \#A, m_i} \sum_{s \hat{=} s_i} p(v, m, s, \omega) \cdot \frac{\Pr(\omega, v, m, s)}{\Pr(v_i, \#A, m_i, s_i)}$$
$$= \frac{\sum_{\omega} \sum_{v \hat{=} v_i} \sum_{m \hat{=} \#A, m_i} \sum_{s \hat{=} s_i} p(v, m, s, \omega) \cdot \Pr(\omega, v, m, s)}{\sum_{\omega} \sum_{v \hat{=} v_i} \sum_{m \hat{=} \#A, m_i} \sum_{s \hat{=} s_i} \Pr(\omega, v, m, s)}$$

Expected payoffs in message stage We are looking for the expectation of payoff of p conditional on message m_i after signal s_i.

$$E(p|m_i,s_i) = \sum_{\omega} \sum_{v \hat{=} v_i} \sum_{m \hat{=} m_i} \sum_{s \hat{=} s_i} p(v,m,s,\omega) \cdot \frac{\Pr(\omega,v,m,s)}{\Pr(m_i,s_i)}$$
$$= \frac{\sum_{\omega} \sum_{v \hat{=} v_i} \sum_{m \hat{=} m_i} \sum_{s \hat{=} s_i} p(v,m,s,\omega) \cdot \Pr(\omega,v,m,s)}{\sum_{\omega} \sum_{v \hat{=} v_i} \sum_{m \hat{=} m_i} \sum_{s \hat{=} s_i} \Pr(\omega,v,m,s)}.$$

Finally, note that $\Pr(m_i,s_i)$ is equivalently defined as

$$\Pr(m_i,s_i) = \sum_{\omega} \Pr(\omega) \cdot \Pr(s_i|\omega) \cdot \Pr(m_i|s_i).$$

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.

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 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.

Table 14: The sequential equilibria that can be represented as limiting logit equilibria in the four treatment

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

	σ		$\tau(a,b,)$			$\tau(b,a,)$			$\tau(b,b,)$			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
Base Rate Fallacy	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: $N = 3, \quad P = 1, \quad K = 15/35, \quad \alpha = 0.6$

	σ		$\tau(a,b,)$			$\tau(b,a,)$			$\tau(b,b,)$			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
Base Rate Fallacy	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: $N = 3, \quad P = 1, \quad K = 10/40, \quad \alpha = 0.6$

	σ		$\tau(a,b,)$			$\tau(b,a,)$			$\tau(b,b,)$			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
Base Rate Fallacy	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: $N = 3, \quad P = 1, \quad K = 15/35, \quad \alpha = 0.6$

	σ		$\tau(a,b,)$			$\tau(b,a,)$			$\tau(b,b,)$			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
Base Rate Fallacy	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

Table 15: Theoretical predictions across treatments

	Messages		Voting											
	$\mu(R)$	$\mu(B)$	$\pi(R,R,0)$	$\pi(B,R,0)$	$\pi(R,R,1)$	$\pi(R,B,1)$	$\pi(B,B,1)$	$\pi(B,R,1)$	$\pi(R,R,2)$	$\pi(R,B,2)$	$\pi(B,B,2)$	$\pi(B,R,2)$	$\pi(R,B,3)$	$\pi(B,B,3)$
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
Base Rate Fallacy	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
Base Rate Fallacy	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
Base Rate Fallacy	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
Base Rate Fallacy	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: $\mu(s)$ is the probability of sending message A given the signal $s \in \{A,B\}$. $\pi(s,m,M)$ is the probability of voting A as a function of one's signal s, message m, and the number M of B messages overall (i.e. in aggregate over all players). The parameters $(\pi_{Lie}, \pi_{\text{Low}}, \pi_{\text{Med}}, \pi_{\text{High}})$ allow adaptation to subjects' behavior, with the theoretical ex-ante hypothesis $\pi_{\text{Low}} < \pi_{\text{Med}} < \pi_{\text{High}}$.

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?

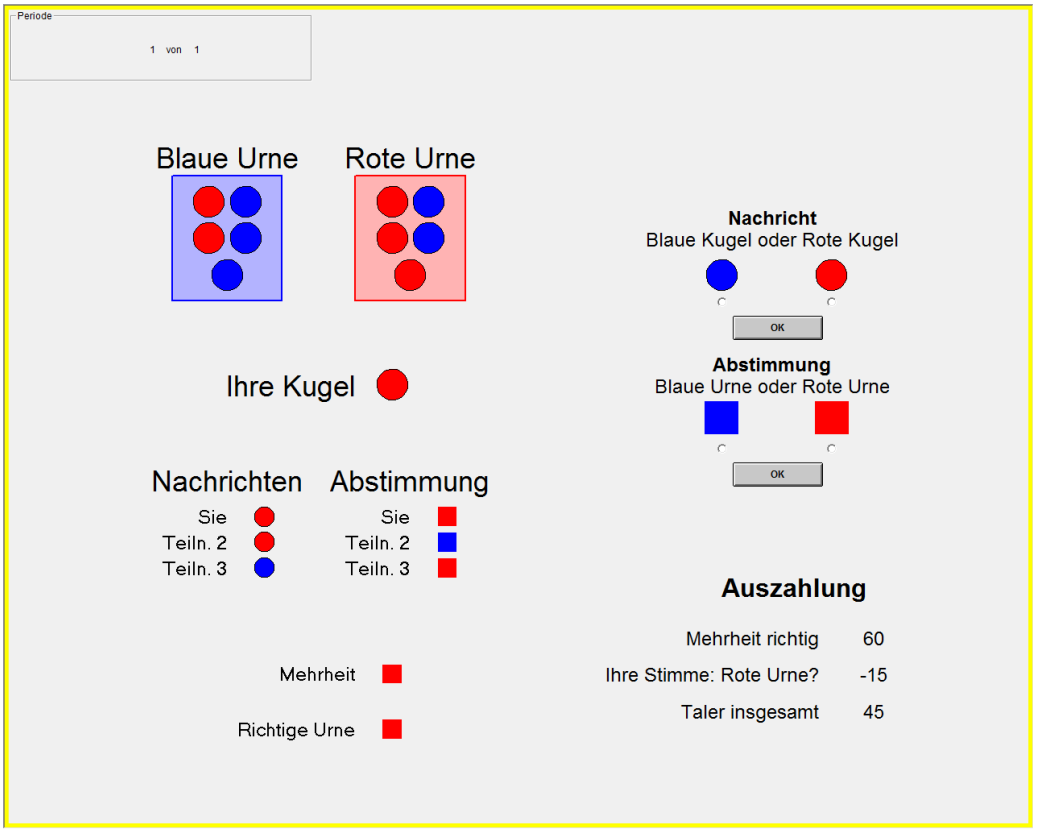
   ☐ 10 Taler ☐ 40 Taler ☐ 50 Taler

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.