Communication and voting in heterogeneous committees: An experimental study

Mark T. Le Quement[†] and Isabel Marcin[‡§]

March 13, 2017

Abstract

We study experimentally the drivers of behavior in committees featuring publicly known diverse preference who engage in voting preceded by one shot cheap talk communication. On the aggregate, we find low lying levels and different preference types using decision rules biased towards the majority heuristic which consists in following the majority of announced signals. Our results are inconsistent with the predictions derived from the standard model as well as models of social preferences and homogeneous naive behavior. Results are instead consistent with the predictions of a model of cognitive heterogeneity, in which a large majority of unsophisticated subjects truth-tells and uses the majority decision heuristic, while a minority of sophisticated agents lies strategically and applies its payoff-maximizing decision rule, albeit with noise.

Keywords: Committees · Voting · Information Aggregation · Cheap Talk · Experiment

JEL Classification: C92 · D72 · D82 · D83

[†]School of Economics, University of East Anglia, Norwich, United Kingdom, NR47TJ. E-mail: M.Le-Quement@uea.ac.uk

[‡]University of Heidelberg, Department of Economics, Bergheimer Str.58, 69115 Heidelberg, Germany, Email: isabel.marcin@awi.uni-heidelberg.de

[§]We thank Christoph Engel, Guillaume Frechette, Sebastian Goerg, Jens Grosser, Oliver Kirchkamp, Sebastian Kube, Michael Kurschilgen, Dimitri Landa, Rebecca Morton, Pedro Robalo, Nicolas Roux, Andrew Schotter, Thomas Palfrey as well as seminar and conference audiences at the ESA Heidelberg, MPI Bonn, NYU Political Economy Workshop, Florida State University Experimental Group, NYU Abu Dhabi Behavioral Political Economy Workshop. Research support from the Max Planck Institute for Research on Collective Goods is gratefully acknowledged. We are grateful to Lars Freund and Anastasiia Niechaieva for research assistance. No IRB approval was required to collect the data.

1 Introduction

Collective decision-making commonly brings together individuals whose preferences are heterogeneous. Examples include parliamentary committees consisting of members of different political parties or boards of directors consisting of different types of stakeholders (public and private stockholders, employees, etc). Collective decision-making often has a common value dimension as members would favour the same decision if the state of the world were known (e.g. whether a defendant is guilty or innocent, whether a reform will lower unemployment, whether a job candidate is competent). Given uncertainty about the state, disagreement may arise because different committee members value the two potential types of error differently (false positive vs. false negative).[1] When heterogeneous preferences are publicly observed, rational and self-interested individuals have incentives to misrepresent their private information, if voting is preceded by a non-binding straw vote (Coughlan, 2000), the latter being equivalent to one round of simultaneous cheap talk. Incomplete information sharing is wasteful from a welfare perspective because pooling of information increases the probability of making a correct decision.

Goeree and Yariv (2011) offer a first experimental study of communication and voting in heterogeneous committees, which constitutes the starting point for our experiment. The authors examine a setting featuring privately known preferences and free-form communication. They find a dominant pattern of behavior consisting of truth-telling followed by voting with the majority of announced signals. The experiment raises two types of questions. The first concerns the robustness of their findings: To what extent is the very high rate of truth-telling found driven by the assumptions of free-form communication and uncertainty about preference types? Both assumptions arguably push towards more truth-telling. Would a simpler communication protocol (e.g. a straw-vote) as well as a more clear cut conflict (i.e. publicly observed heterogeneous preference types) give rise to more strategic communication?

The second type of question raised by Goeree and Yariv (2011) concerns the underlying drivers of the behavior that they document. The authors explain that "groups make choices that are consistent with the welfare maximizing decisions given the available aggregate information in the group" (see p. 895). This explanation implicitly alludes to subjects exhibiting social preferences, but the authors take no explicit stance on the issue and make no attempt to test

[1] For example, there is a large body of empirical evidence showing that jury members hold heterogeneous preferences regarding the possibility of convicting an innocent defendant vs. acquitting a guilty defendant. Different preferences may be rooted in political attitudes, demographic characteristics, personality traits, etc. In the psychological literature on judicial decision-making, empirical studies examine the effects of demographic and personal characteristics on behavior in jury deliberation (for reviews see MacCoun, 1989; Pennington and Hastie, 1990; Devine et al., 2001; Sommers and Ellsworth, 2003).

alternative models. One can envisage two potential deviations from standard assumptions that may positively affect the amount of truth-telling observed, namely social preferences and cognitive constraints. If, for instance, individuals' objective function is to maximize joint payoffs instead of individual payoffs, lying incentives are eliminated. Alternatively, agents may face cognitive constraints that make them unable to communicate strategically or to identify their payoff-maximizing decision rule, leading them to use simple heuristics (e.g. truth-telling, following the majority of signals).

Our experiment is aimed at addressing the above two classes of questions. We examine a three-person deliberative jury game (Feddersen and Pesendorfer, 1998; Coughlan, 2000) with two publicly known preference types and majority rule and explicitly test not only the standard model but also a set of alternatives. The assumptions of (1) known heterogeneous preferences as well as (2) communication in the form of straw votes are both empirically and theoretically motivated. Preferences often correlate with observable characteristics and committee members often have relevant information about each other. Theoretically, the setups exhibit very different predictions. Full pooling is essentially more difficult to achieve with publicly known heterogeneous preferences, so that we expect strategic lying to be more salient. As to our focus on straw votes (pre-defined messages), these would appear to minimize the scope for the emergence of social preferences through communicative interaction, as compared to free form chats, thus in principle offering more potential for strategic interaction.

Our experimental design allows us to investigate the underlying drivers of behavior, i.e., whether social preferences or cognitive constraints significantly affect communication (or the potential absence of strategic communication). Our design varies (a) the information provision protocol (public signals vs. private signals) and (b) the committee composition (homogenous vs. heterogeneous), thus yielding four treatments. Consequently, we can observe voting with and without prior communication keeping constant the number of signals that are available in the committee. This comparison is novel in the literature, which has focused on comparing private voting and voting after communication of private signals (Guarnaschelli et al., 2000; Goeree and Yariv, 2011).

We derive testable predictions from three models, the "standard" model of own payoff-maximizing strategic agents (Feddersen and Pesendorfer, 1998; Coughlan, 2000; Le Quement and Yokeeswaran, 2015) and two behavioral alternatives. The first alternative is a social preference model of joint payoff maximization. The second alternative is a naïve voter model where all individuals truth-tell and use a decision heuristic which specifies voting in line with the majority of announced signals (*majority heuristic*). This second model is in the spirit of Condorcet (1785), who assumes non-strategic jury members who simply favour the alternative that is the

most likely to be correct.

On aggregate, our results reject the predictions derived from the standard model and both behavioral alternatives: (a) We find no evidence of babbling as predicted by our equilibrium for the standard model; (b) given public signals, the preference profile of the committee (heterogeneous vs. homogenous) does not affect individuals' voting behavior as predicted by the social preference model; (c) the two different preference types do not use the same decision rule as predicted by the naïve voter model.

In a second, more preliminary and explorative part of our analysis, we disaggregate results across individuals and find significant heterogeneity. We introduce a model of level-k thinking (see for instance Stahl and Wilson, 1994, 1995; Nagel, 1995; Crawford and Iriberri, 2007) and noisy behavior. Level-0 agents behave as in the naïve voter model. Level-1 agents best respond in a noisy fashion to the assumption that all others are level-0 agents. They lie when they hold a signal that is contrary to their preference bias and apply their payoff-maximizing decision rule at the voting stage. Our findings suggest that subjects can be meaningfully assigned to these two general level-0 and 1 categories. The vast majority of subjects (82%) consistently truth-tells and uses its type-specific optimal decision rule only very seldom, in only 24% of all cases, reverting otherwise to the majority-heuristic. In contrast, 18% of subjects consistently lie after a signal contrary to their preference bias and use their type-specific optimal decision-rule on average 60% of the time. Consistent lying thus significantly associates with applying the type-specific payoff-maximizing decision rule.

Summarizing, our analysis yields two main findings. First, at an aggregate level, dominant truth-telling is not simply an artefact of free-form communication and unknown preference types. Second, aggregate results cover up heterogeneous underlying behavior that is consistent with the presence of different cognitive types. To the best of our knowledge, this is the first experimental study to indicate the relevance of the level-k thinking model to committee voting, in particular the variant with communication. Generally speaking, the finding is in line with pre-existing experimental findings on level-k play in cheap talk games (see Cai and Wang, 2006; Wang et al., 2010; Kawagoe T, 2009).

We proceed as follows. Section 2 presents a brief overview of the literature. Section 3 presents our experimental design and Section 4 the theoretical predictions. Section 5 analyzes aggregate behavior with an eye to testing the respective predictions. Section 6 focuses on heterogeneity in behavior, presents the cognitive heterogeneity model and analyzes its predictions. Section 7 concludes.

2 Related Literature

Building on Condorcet's seminal essays on voting (see Condorcet, 1785), a theoretical literature that models voting as information aggregation has blossomed over the last two decades. Early contributions (Austen-Smith and Banks, 1996; Feddersen and Pesendorfer, 1998) study private voting (see also Gerardi, 2000; Martinelli, 2006; Meirowitz, 2007; Persico, 2004; Feddersen and Pesendorfer, 1996). Key findings have been confirmed and qualified experimentally in Guarnaschelli et al. (2000), Esponda and Vespa (2014), Grosser and Seebauer (2016), Battaglini et al. (2008), and Battaglini et al. (2010).

A set of newer contributions study the case of voting preceded by communication and have focused on the truthful-sincere equilibrium. A milestone is the negative result obtained by Coughlan (2000) for the case of publicly known heterogeneous preference types: If full truth-telling leads to disagreement, then there exists no truthful-sincere equilibrium. Le Quement and Yokeeswaran (2015) provide an equilibrium prediction for such committees under unanimity rule. Deimen et al. (2015) offer a complementary analysis that assumes conditionally correlated signals. A parallel research agenda has studied the extent to which uncertainty about preference types affects the possibility of communication (Austen-Smith and Feddersen, 2006; Meirowitz, 2007; Van Weelden, 2008; Le Quement, 2013).

Voting with communication has also been examined experimentally.[2] Guarnaschelli et al. (2000) study a homogeneous jury and find, in contradiction with the intuitive prediction of full truth-telling, a 5% lying rate and skepticism towards information provided by others. Goeree and Yariv (2011) study the case of privately known (and potentially heterogeneous) preference types with free-form communication. The authors' confirm the theoretical prediction formulated in Gerardi and Yariv (2007), namely that all voting rules are equivalent given unrestricted communication. Furthermore, they find that subjects on average follow a simple heuristic consisting in truth-telling combined with voting according to the majority of announced signals.

In seeking to identify the drivers of behavior in the communication and voting game, we draw on ideas taken from two strands of literature, namely social preferences and bounded rationality, each of which proposes models that have been successfully applied to a variety of games. The social preferences literature features outcome-based models focusing on inequity aversion or taste for efficiency as well as intention-based models that analyze the role of reciprocity, kindness, etc (see for example Rabin (1993); Fehr and Schmidt (1999); Bolton and Ockenfels (2000); Charness and Rabin (2002)). The social preference model that we use is an outcome-based model and

[2]Our focus, as well as that of the here reviewed literature, is on deliberation as information aggregation. We refer to Hafer and Landa (2007) and Dickson et al. (2008) for theoretical and experimental work on deliberation modeled as a rational and strategic process of self-discovery.

assumes that agents care about the sum of committee members' payoffs, reflecting both a taste for efficiency and inequity aversion. As to the literature on bounded rationality, we use ideas from two sub-strands, namely the level-k reasoning model and the quantal response model. The former (see Stahl and Wilson, 1994, 1995; Nagel, 1995; Crawford and Iriberri, 2007) assumes agents exhibiting different depths of reasoning.[3] The model has successfully been used to describe behavior in cheap talk games (Cai and Wang, 2006; Wang et al., 2010; Kawagoe T, 2009). The quantal response model (McKelvey and Palfrey, 1995, 1998) assumes that agents act noisily in response to correct equilibrium beliefs. The simple model of bounded rationality that we propose in the last part of our analysis combines elements of these two models.

3 Experimental Design

We here describe the treatments and the experimental procedure.

3.1 Treatments

Our main treatment is a Condorcet Jury Game with private information and heterogeneous preferences. A committee, composed of three subjects, has to choose between two alternatives by majority vote. If the state of the world were known, each preference type would favor choosing the alternative that matches the true state of the world. However, different preference types value the two potential errors differently. The state of the world is not observable, but takes one of two possible values, both being ex ante equally probable. Specifically, the state is the (red or blue) jar selected by nature and the decision is either *red* or *blue*.[4]

The timing of the game is as follows. There is an information stage (stage 1) at which information regarding the color of the jar is received and exchanged. At stage 2, each subject casts a vote from the set $\{red, blue\}$ and a collective decision is made. In stage 3, subjects observe the number of votes for each jar, the committee decision, the jar selected by nature as well as their payoffs.

We vary the main treatment on two dimensions: (a) the preferences of subjects and (b) the information protocol, see Table 1 below.

Preferences (Heterogeneous vs. Homogenous): In heterogeneous (*Het*) committees there are two possible preference types, *red* or *blue*, whose payoffs depend on the group decision

[3] See also Goeree and Holt (2004) for a related model of noisy introspection.

[4] We follow Guarnaschelli et al. (2000) and Goeree and Yariv (2011) in adopting a neutral description. In the jury interpretation, the group chooses between convicting and acquitting a defendant who is either guilty or innocent.

Table 1: Overview of treatments

		Preferences	
		Homogeneous	Heterogeneous
Information	Private signals	Private-Hom	Private-Het
	Public signals	Public-Hom	Public-Het

and the realized jar (see Table 2). As can be seen from the table, red (blue) types are biased towards the red (blue) jar. If agents are risk neutral and self-interested, this payoff specification is equivalent to the preference specification introduced in Feddersen and Pesendorfer (1998) and Coughlan (2000).[5] Committee composition is common knowledge at the start of the game. In each committee there are either two *red* types and one *blue* type or vice-versa.[6] In contrast, homogeneous (*Hom*) committees consist only of one preference type, either all subjects are *red* or *blue*.

Table 2: Payoff structure

		True Jar		True Jar	
		Blue Jar	Red Jar	Blue Jar	Red Jar
Group Decision	Red	10	40	10	160
	Blue	160	10	40	10
		Blue Type		Red Type	

Information (Private vs. Public): In *private* treatments, information is transmitted in two stages. In substage 1.a, each agent privately observes a signal. A signal takes the form of a red or blue ball randomly drawn with replacement from the realized jar. The blue (red) jar contains 7 (3) blue balls and 3 (7) red balls.[7] In the subsequent substage 1.b, each agent picks a simultaneously observed public message from the set $\{red, blue\}$.[8] In other words, the *private*

[5] In those models, a juror's payoff is determined by a commonly known parameter $q \in (0,1)$. He obtains payoff $-q$ (resp. $-(1-q)$) if the chosen jar is red (blue) while the realized jar blue (red). Payoffs from choosing the correct jar are normalized to 0. We exclude negative payoffs by applying a positive transformation to the original ones. Payoffs in Table 2 are equivalent to $q_{blue} = \frac{5}{6}$ for blue types and $q_{red} = \frac{1}{6}$ for red types in the original specification.

[6] A subject whose preference type is (not) shared by some (any) other subject is called a majority (minority) subject.

[7] Formally, a signal s is an independent Bernoulli trial from a state-dependent distribution with $P(s = red \mid red) = P(s = blue \mid blue) = p = 0.7$, while $P(s = blue \mid red) = P(s = red \mid blue) = 1 - p = 0.3$.

[8] Note that a subject does not have the possibility to refrain from sending a message, as in the original Coughlan

treatments feature a round of simultaneous cheap talk communication. Messages are shown with an indication of the subject's preference type, but without a player identifier. In so-called *public* treatments, information comes in the form of three conditionally i.i.d. public signals, conditional on the state, which is equivalent to forced sincere communication.

We introduce some simplifying notation. First, we shall repeatedly be referring to the *observed signal profile* of a subject in stage 1. In *private* treatments, it corresponds to a subject's own signal combined with the two signals announced by others. In *public* treatments, it corresponds to the three public signals observed. Second, we call a red signal held by a red type a *conform* signal and a blue signal held by a red type a *contrary* signal. Equivalently, we call decision *red* (*blue*) the *conform (contrary)* decision for a red type (and vice versa for blue types). As can be seen in Table 2, payoffs are symmetric across red and blue types. Thus, we should expect identical behavior by red and blue types at symmetric information sets. Consider an outcome given by a profile of three signals in a public treatment combined with a decision. Construct the symmetric outcome, which is obtained by replacing any blue (red) signal by a red (blue) signal as well as reversing the decision. The expected payoff of a red (blue) type given the first outcome is the same as that of the blue (red) type given the second outcome.

3.2 Experimental Procedure

We use a between-subjects design. Each session contains the following parts: (1) treatment, (2) strategic communication test (SCT) (only *private*), (3) individual decision test (IDT), (4) lying aversion test and (5) social value orientation test. Payoffs from each of the post-experimental tests are learned after the last test. While (4) and (5) are standard tests adopted from the literature, (2) and (3) are novel. The SCT test evaluates subjects' ability to communicate strategically. It is only taken by subjects in the private treatments (as these involve communication). The IDT measures subjects' decision rule in an individual decision task and thus excludes effects related to beliefs about others' behavior or social preferences. See Appendix A.2 for a description of the post-experimental tests (2)-(5).

At the start of each treatment, subjects are randomly assigned a preference type and a matching group of 6 subjects.[9] In each period, two three-subject committees are randomly formed. An equal number of subjects is assigned to each preference type. In *Hom* treatments, each matching group contains either only blue or only red types. In *Het* treatments, each matching group contains three blue and three red types. The game is played repeatedly over 20 rounds with random rematching within each matching group. In *Het* treatments, each subject

(2000) setup.

[9]Subjects are not informed about the size of the matching group.

is thus very likely to experience multiple rounds in minority and in majority.

The experiment was conducted in the BonnEconLab in February and March 2015. It was programmed and conducted with the software z-Tree (Fischbacher, 2007) and organized with the software hroot (Bock et al., 2014). A total of 384 University of Bonn students from various disciplines (15% with an economics major) participated in 16 sessions (each of 24 subjects). 96 subjects participated in each treatment, yielding 16 independent matching groups per treatment. Subjects received written instructions which were read out loud by the experimenter (see Appendix A.4 for an English transcript of the original German instructions). To familiarize subjects with the game and ascertain that they understood it fully, we asked control questions that had to be answered correctly. Subjects were given the opportunity to privately ask questions. The amounts earned from the experiment were exchanged at a rate of 150 ECU = 1 Euro. Subjects received the payment from all 20 rounds, which averaged 10.50 Euros and ranged from 5.50 Euros to 16.50 Euros. Subjects additionally earned an average of 4.68 Euros in the post-experimental tests. On average, one session lasted 65 minutes (40 minutes jury experiment and 25 min post-tests). 58.6 % of subjects were female and average age was 22.6 years.

4 Theoretical Predictions

This section introduces the standard model, the social preference model of joint-payoff maximization and the naïve voter model. The first two models assume rational and risk-neutral agents while they differ on the assumed preferences, i.e. agents maximize own payoffs in the standard model and joint payoffs in the social preference model. We focus on equilibria in symmetric strategies, in which agents with identical payoff functions use the same strategy. For each treatment, we obtain theoretical predictions for each of the three models and subsequently derive a set of testable hypotheses concerning differences in outcomes across treatments.

4.1 Standard Model

The standard model is analyzed in Feddersen and Pesendorfer (1998) and Coughlan (2000). It assumes that agents only maximize own expected payoffs and are risk-neutral. Given the payoffs in Table 2, an agent favors the conform decision (the latter being the red decision given that he is red-biased and the blue decision when blue-biased) if the conform jar has a conditional probability of at least $\frac{1}{6} \approx 0.167$. The conditional probability of a conform jar after 0, 1, 2 and 3 conditionally i.i.d conform signals of quality $p = .7$ is given by respectively .07, .3, .7 and .93. The payoff-maximizing decision rule of each preference type is thus to choose the conform decision if at least one of the three signals is conform. We denote by $\Lambda(x)$ the decision rule specifying the

following probabilities of picking the conform decision after r conform signals:

$$[\Lambda(x)](r) = \begin{cases} 0 & \text{if } r = 0, \\ x & \text{if } r = 1, \\ 1 & \text{if } r \geq 2. \end{cases}$$

The rule $\Lambda(1)$ is thus the payoff-maximizing or optimal decision rule of each type. For a given preference type $j \in \{red, blue\}$, we call *conform-based decision rule* a decision rule that maps from the number of j-signals observed to the probability of choosing decision j. We call *j-based decision rule* a decision rule that maps from the number of j-signals observed to the probability of choosing decision j.

The impossibility result obtained by Coughlan (2000) implies that if the committee contains at least one blue-biased and one red-biased agent, there exists no equilibrium in which all agents truth-tell and vote sincerely. To understand the result, assume that the committee contains a simple majority of blue-biased agents. The decision rule applied in the above putative equilibrium is the optimal decision rule of blue-biased agents, i.e. choose red only if three red signals are observed. At the communication stage, the red-biased agent acts under the assumption that his announcement is pivotal (i.e. affects the final outcome) and thus infers that the two other agents hold a red signal. This in turn implies that he favors a red decision. If he holds a blue signal, he thus deviates to announcing a red signal. Our equilibrium prediction for each of the treatments is given below. For *private* treatments, we restrict ourselves to equilibria featuring truth-telling by the numerically dominant type and sincere voting by both types. As the numerically dominant type can always get its way at the voting stage given majority rule, truth-telling by this type appears salient. We obtain the following equilibrium characterization.

Proposition 1.

a. Private-Het: Majority agents truth-tell while the minority agent babbles. Majority agents condition their vote only on majority agents' signals. They vote for the conform decision unless they jointly hold two contrary signals. The minority agent conditions his vote on all members' signals and applies $\Lambda(1)$ *to the observed signal profile.*

b. Private-Hom: All agents truth-tell. All agents apply $\Lambda(1)$ *to the observed signal profile.*

c. Public-Het: All agents apply $\Lambda(1)$ *to the observed signal profile.*

d. Public-Hom: All agents apply $\Lambda(1)$ *to the observed signal profile.*

The intuition for our prediction for the *Private-Het* treatment is as follows. At the voting stage, the optimal decision rule of the majority preference type conditional on two signals is implemented. This involves choosing the conform decision unless the two signals are contrary.

The minority agent is never pivotal at the voting stage and is thus indifferent between both voting decisions. At the communication stage, a majority agent recognizes that his optimal decision rule is implemented given the publicly pooled information. A majority agent's announcement is pivotal at a unique signal constellation which encourages truth-telling. Assume that red-biased agents are the majority and consider a red-biased agent i. The unique pivotal scenario is when he holds a red signal and others hold blue signals. Announcing a red (blue) signal leads to a red (blue) decision. Indeed, while the voting decision of agent i and the minority agent is independent of i's announcement (the first votes red, the other one blue), the other red-biased agent only votes red if i announces red. Clearly, i prefers to truth-tell. On the other hand, the communication incentives of a minority agent are trivial. Given that his announcement is ignored, he is indifferent between all messages and accordingly has no incentive to deviate from babbling. As to *Private-Hom*, note that truth-telling is trivially incentive compatible as an agent knows that his optimal decision rule is implemented at the decision stage given pooled information.

We derive three treatment hypotheses from the above proposition concerning (a) communication, (b) use of information and (c) voting behavior.

Set of Hypotheses 1.

a. Communication: In private treatments, communication by minority subjects in heterogeneous committees is less informative than (i) communication by majority subjects in heterogeneous committees and (ii) communication by subjects in homogenous committees.

b. Voting in private treatments: In Private-Het, majority subjects condition their vote less on the announcement of minority subjects than on that of majority subjects.

c. Voting in public treatments: In public treatments, the frequency of a conform vote given an observed signal profile containing one conform signal is the same in homogeneous and heterogeneous committees.

The hypothesis regarding voting behavior that appears in c. is formulated only for *public* treatments because these by definition exclude any potential skepticism towards information arising as a consequence of communication. These treatments thus provide clean evidence of how subjects decide on the basis of unambiguously trustworthy public information. We focus on behavior given a single conform signal because we expect most of the variation in behavior (across subjects or treatments) to arise at this particular information set.

4.2 Social Preference Model

In this model we assume that agents maximize the sum of committee members' individual type-specific payoffs. Agents thus behave as if they all shared the same payoff function given by the

average payoff function. In a committee with two (one) blue agents and one (two) red agent, this implies that agents require a conditional probability of the red jar of approximately 0.61 (.39) in order to favor the red decision. Accordingly, the optimal decision rule conditional on three signals is to vote in line with the majority of signals ($\Lambda(0)$), whatever one's preference type. Our selected equilibrium prediction for this model features full truth-telling for any committee composition. Full truth-telling appears very intuitive given that all jurors share de facto identical goals and thus understand that they have no reason to distrust each other. Our focus on equilibria featuring the maximal achievable amount of information pooling is furthermore in line with our analysis of the standard model.

Proposition 2.

a. Private-Het: All agents truth-tell and apply $\Lambda(0)$ to the observed signal profile.

b. Private-Hom: All agents truth-tell and apply $\Lambda(1)$ to the observed signal profile.

c. Public-Het: All agents apply $\Lambda(0)$ to the observed signal profile.

d. Public-Hom: All agents apply $\Lambda(1)$ to the observed signal profile.

Though committee composition does not affect communication in the above predictions, it however affects the implemented decision rule. While heterogeneous committees vote in line with the majority of signals, homogenous committees implement the type-specific decision rule $\Lambda(1)$. Note that our model corresponds to the extreme point of a continuum of models in which a parameter (say $\alpha \in [0, 1]$) measures the degree of altruism of agents. Agents maximize a function given by α times their individual payoff and $1-\alpha$ times the total committee payoff. We set $\alpha = 0$ for simplicity of exposition, but our predictions for all the treatments would still hold for α small enough.[10] We derive the following set of hypotheses from the above proposition.

Set of Hypotheses 2.

a. Communication: In private treatments, communication is equally informative (i) in homogeneous and heterogeneous committees, and (ii) across majority and minority subjects.

b. Voting in private treatments: In Private-Het, subjects condition their vote equally on all signals featured in the observed signal profile.

c. Voting in public treatments: In public treatments, subjects apply different decision rules depending on whether they are in a homogeneous or heterogeneous committee. The frequency of a conform vote given an observed signal profile containing one conform signal is higher in homogeneous committees than in heterogeneous committees.

[10] Namely, $\alpha \leq 0.699$ for minority *Private-Het* subjects and $\alpha \leq 0.402$ for majority *Private-Het* subjects. While the specific utility function assumed allows us to generate point predictions, other forms of social preferences, as inequity aversion (Fehr and Schmidt, 1999) or a taste for efficiency (Charness and Rabin, 2002) would also predict treatment differences.

4.3 Naïve Voter model

In his seminal essay, Condorcet (1785) assumed that all individuals have the same objective of making a correct decision, ignore the strategic aspects of committee-decision making and simply vote as if they were the only voter (i.e. sincerely). Translating this idea into our setting means that agents truth-tell and vote in line with the majority of announced signals. They thus choose the alternative that is most likely to be true without taking into account their expected payoffs. This naïve voter model is also in line with the behavior of subjects in treatments with communication in the experiment of Goeree and Yariv (2011).

Proposition 3.

a. Private-Het: All agents truth-tell and apply $\Lambda(1)$ *to the observed signal profile.*

b. Private-Hom: All agents truth-tell and apply $\Lambda(1)$ *to the observed signal profile.*

c. Public-Het: All agents apply $\Lambda(1)$ *to the observed signal profile.*

d. Public-Hom: All agents apply $\Lambda(1)$ *to the observed signal profile.*

Set of Hypotheses 3.

a. Communication: In private treatments, communication is equally informative (i) in homogeneous and heterogeneous committees, and (ii) across majority and minority subjects.

b. Voting in private treatments: In Private-Het, subjects condition their vote equally on all signals featured in the observed signal profile.

c. Voting in public treatments: In public treatments, subjects apply the same red-based decision rule independent of (i) whether they are in a homogeneous or heterogeneous committee and (ii) whether their preference type is blue or red.

5 Results: Aggregate behavior

In what follows, we analyze aggregate communication and voting behavior and test the hypotheses formulated in our theoretical predictions section. We pool red and blue types.[11]

5.1 Communication

Table 3 shows average lying rates based on individual averages conditional on the signal received for *Private-Hom* subjects, minority *Private-Het* subjects and majority *Private-Hom* subjects.

[11] As already noted earlier, this should be unproblematic given the symmetry of payoffs across types. This is confirmed by statistical analysis. For each type of signal held (conform or contrary) and each possible committee position (i.e. *Private-Het* majority, *Private-Het* minority or *Private-Hom*), we do not find significant differences across preference types in terms of voting and communication decisions (two-sided Mann-Whitney rank-sum test) (MW test in what follows).

The lying rate after a conform signal is approximately 0 for all three types of subjects. On the other hand, the lying rate after a contrary signal is substantially larger for all three types, though it remains low in absolute terms. In the table below, recall that in *Hom* treatments, a subject is by definition always a majority subject.[12]

Table 3: Lying rates in *private* treatments in %

	Private-Hom		Private-Het	
Signal	lying rate	n	lying rate	n
contrary in minority			21.9 (34.44)	324
contrary in majority	10.2 (22.52)	1019	14.9 (27.11)	635
conform in minority			1.0 (5.94)	316
conform in majority	0.7 (4.8)	901	0.5 (2.86)	645

Notes: Standard deviation are denoted in brackets. n denotes the number of observations in each respective information set.

We find no evidence of babbling by minority *Private-Het* subjects. Recall that babbling by a given subject requires the latter to use the same distribution over messages after each signal that he might hold, the choice of the distribution being arbitrary. Given that minority *Private-Het* subjects have a lying rate of almost 0 after conform signals, babbling would imply that they always lie after a contrary signal. A one-sided t-test clearly rejects this conjecture ($p < 0.001$). For this and all following tests, we use matching groups as independent units of observation. We find marginally significant evidence that minority *Private-Het* subjects lie more after contrary signals than majority *Private-Het* subjects (one-sided Wilcoxon signed-rank, WX test, $p = 0.098$) and significant evidence that minority *Private-Het* subjects lie more than *Private-Hom* subjects (one-sided Mann-Whitney, MW test, $p = 0.02$). The increased lying rate of minority subjects is in line with the idea of the unilaterial deviation scenario arising in the hypothetical truthful-sincere equilibrium analyzed by Coughlan (2000), i.e. minority types lie to majority types hoping that their message is taken at face value, which in turn would bend the majority types' decision rule towards the one of the minority type.

To measure whether subjects lie more over the course of periods, we run a regression with the dummy variable *lie after a contrary signal* as dependent variable (see Table A2 in Appendix). We find that over time subjects lie significantly more, however, the effect is small. In this analysis, we also test the impact of lying aversion (as measured in the post-experiment lying aversion test)

[12]Lying rates of subjects in *Private-Hom* are shown in the row *majority*.

on lying behavior in the treatments. Lying aversion has little explanatory power overall but strongly correlates with the lying behavior of the subsample of subjects who lied at least once.[13]

Summarizing, behavior at the communication stage thus yields mixed results. There is no evidence of babbling by minority subjects as specified by our equilibrium prediction for the standard model. On the other hand, minority subjects in *Private-Het* lie significantly more than subjects in *Private-Hom*, which is at odds with both behavioral models.

Result 1.

Subjects to a large extent truth-tell. However, there is more lying by minority subjects after contrary signals than after conform signals and marginally more truth-telling in homogenous committees than in heterogeneous committees.

5.2 Voting in private treatments

The standard model predicts that majority subjects in heterogenous committees ignore the information provided by minority types. Both alternative models, however, predict that subjects conditional equally on all announced signals (as everybody is predicted to truth-tell). We therefore focus on majority *Private-Het* subjects and analyze the extent to which they condition their voting decision on the announcement of minority *Private-Het* subjects. Table 4 shows a majority type's frequency of choosing the conform decision as a function of his own signal (a conform signal takes the value of 1 and a contrary signal takes the value of 0) and the announcement of the two remaining subjects, one majority type and one minority type. Choice frequencies show that the information provided by the minority type is influential. To see that, compare choice frequencies in cases that differ only according to the message announced by the minority type: 1 vs 3, 2 vs 4, 5 vs 7 and 6 vs 8. Choice frequencies furthermore show that minority type announcements are approximately as influential as majority type announcements (compare cases 2 vs 3 and 6 vs 7).

The above findings are confirmed by statistical analysis. We perform two comparisons that each hold the observed signal profile constant. The first comparison involves cases 2 and 3, for which our respective theoretical predictions, as derived from the standard model, differ. Both cases involve an observed signal profile containing only one conform signal. When the conform message is sent by the other majority type (Case 2), the prediction for the majority type is to vote conform. When the same message is however sent by the minority type (Case 3), the prediction is a contrary vote. We find that the frequencies of a conform decision do not

[13] In the appendix we also examine behavior in the SCT that was designed to test subjects' ability to communicate strategically. It provides no clear insights as to whether subjects understood the incentive of strategic lying. Many subjects continue acting in the SCT as in the treatment, rendering the SCT results little informative. See Appendix A.3.4 for a discussion.

Table 4: Voting behavior by majority types in Private-Het

Case	n	Observed signal profile			Voting behavior		WX
		own	other majority	other minority	predicted	actual	p-value
1	188	0	0	0	0	0.11	
2	219	0	1	0	1	0.19	0.2^1
3	106	0	0	1	0	0.22	
4	122	0	1	1	1	0.96	
5	161	1	0	0	1	0.45	
6	184	1	1	0	1	0.97	0.046^2
7	88	1	0	1	1	1.0	
8	212	1	1	1	1	1.0	

Notes: n denotes the number of observations of each case. The observed signal profile includes the majority type's *own* signal, the message by the other *majority* type and the message by the *minority* type. The voting behavior shows the *predicted* frequency to vote conform (according to the standard model) and the *actual* frequency to vote conform. WX is a Wilcoxon signed-rank test ([1]one-sided,[2]two-sided). The unit of independent observation is the matching group.

differ between cases 2 and 3 (one-sided WX test, $p = 0.2$). The second comparison involves cases 6 and 7, for which the standard model predicts the same frequency of conform votes. In cases 6 and 7, the majority type holds a conform signal, which implies that he favors a conform decision independently of the signals announced by others. A two-sided WX test rejects ($p = 0.046$) the hypothesis that the frequency of a conform decision in case 6 is equal to that in case 7. A minority type's announcement is actually slightly more influential than that of a majority type. A potential explanation is that a minority type announcing a signal that contradicts his bias (e.g. a blue-biased subject announcing a red signal) is naturally perceived as credible. Summarizing, information sent by minority subjects is thus not disregarded. This result contradicts the prediction of the standard model, but is in line with both behavioral models.

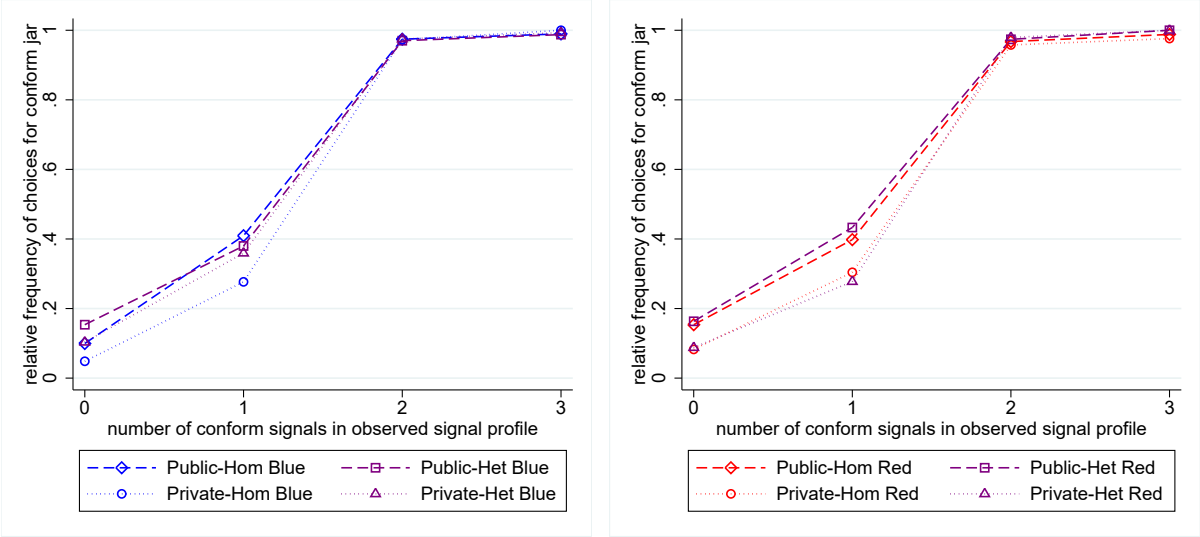
Result 2.

In heterogeneous committees majority subjects condition their vote on the announcement of the minority subject. Majority subjects condition their vote on the announcement of the minority subject approximately as much as on that of a majority subject.

5.3 Voting in public treatments

Figures 1a and 1b show, for each treatment, the frequencies of votes for the conform jar of each preference type as a function of the number of conform signals in the observed signal profile. For all treatments and preference types, subjects vote conform with a probability that is clearly smaller than one (around 0.35) given a unique conform signal, which is also confirmed by statistical tests. A t-test rejects that the frequency of a conform vote given one conform signal is equal to 1 in the *public* treatments (one-sided t-test, $p < 0.001$). Average decision rules thus exhibit what could be termed a reversal to the middle. This is also reflected in the results of the IDT, which excludes any role of strategic interaction or social preferences. Only 27.34% of all subjects apply their payoff-maximizing decision rule, but 70.31% apply the majority heuristic. For further information on the IDT see A.3.5.

Fig. 1: Frequencies of choices for the conform jar



(a) Blue types

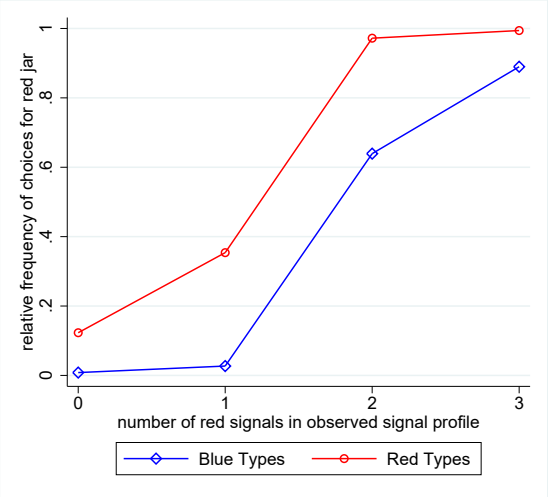
(b) Red types

The social preference model predicts for the heterogeneous committee in the *public* treatment that the frequency of a conform vote given one conform signal is equal to 0, which is also rejected by a one-sided t-test ($p < 0.001$). In other words, subjects do not apply $\Lambda(0)$ to the observed signal profile either. More importantly, the frequency of a conform decision given one conform signal does not differ between homogenous and heterogeneous committees in the *public* treatments (two-sided MW, $p = 0.89$), which means that voting decisions do not depend on the group composition, which the social preference model would have predicted.

The naïve voter model predicts that subjects vote with the majority of announced signals, i.e. vote red if there are at least two red signals (and vice-versa for blue), *independent* of their

own preference type. To test this prediction, we look at the voting decision *conditional on red signals* in the observed signal profile (see Figure 2). Both after one and after two red signals, the probability of a red vote by a red type is much larger than the corresponding probability for a blue type (two-sided MW, $p < 0.01$, in each treatment).

Fig. 2: Frequencies of choices for the red jar across types



Notes: This figure displays voting behavior for red and blue types pooled over treatments.

Result 3.

Voting decisions in public treatments do not depend on the group composition. Both types apply roughly the same conform-based decision rule given by $\Lambda(.4)$. *In other words, the red-based decision rule applied by blue types is significantly different from the red-based decision rule applied by red types.*

We briefly add a comment on risk aversion. If subjects had very concave utility functions and were thus very risk averse, the utility maximizing decision rule would be $\Lambda(0)$ given three signals. To test whether risk attitudes influenced behavior, we run a regression for the *public* treatments where the dependent variable is a dummy equal 1 if the subject votes for the conform decision and 0 otherwise. The coefficient for risk aversion is marginally significant ($p \leq 0.10$) and small in size, in contrast to the coefficients for the IDT threshold and the dummies for the number of conform messages. We therefore conclude that risk aversion had a negligible impact on behavior (see Appendix A.3.2 and regression (1) in Table A3). We also controlled for learning effects in the regression and find that subjects voted more often for the conform jar over the course of time, albeit the effect is small in size.

5.4 Welfare

Table 5 shows average payoffs per treatment (based on matching group averages). Payoff differences between *Private-Het* and *Public-Het* are marginally significant (MW, $p < 0.097$), while there are no significant differences between *Private-Hom* and *Public-Hom*. Transiting from the *public* to the *private* treatments thus only causes a significant loss of information in the case of heterogeneous committees. Note that we focus on payoffs as welfare criterion since minimizing the number of wrong outcomes is the welfare-maximizing rule only in heterogeneous and thus not in homogenous committees (see Table A1 in Appendix on the share of wrong decisions per treatment and type).

Table 5: Average payoffs per treatment

Public-Hom	Public-Het	Private-Hom	Private-Het
81.39 (10.03)	79.92 (5.74)	78.86 (11.26)	75.73 (7.54)

Notes: Standard deviation are denoted in brackets.

5.5 Summarizing findings

Our equilibrium prediction for the standard model is not borne out by the data. We find no evidence of babbling by minority subjects in *Private-Het*. Second, majority subjects condition their vote roughly as much on minority subjects' announcements as on those of majority subjects. Third, in *public* treatments, each type's average decision rule significantly differs from the payoff-maximizing decision rule. Instead, each preference type's average decision rule is heavily skewed towards the *majority heuristic* that consists in voting for the decision indicated by the majority of signals. The prediction derived for the social preference model is also contradicted. In the *public* treatments, we observe no significant shift in decision rules between heterogeneous and homogenous committees, which indicates that committee composition does not significantly affect subjects' payoffs. Finally, the prediction of the naïve voter model is also not matched by the data. Blue and red types apply significantly different red-based decision rules. Also, minority subjects lie slightly more than majority subjects in heterogeneous committees (albeit lying rates remain very low in absolute terms), and they lie more after contrary signals than after conform signals.

6 Heterogeneity in behavior

Our previous section on aggregate behavior documents a low lying rate and an "intermediate" decision rule that lies in between the type-specific optimal decision rule and the majority heuristic. In the following section, we analyze whether the observed aggregate behavior reflects homogeneous individual behavior or rather covers individual heterogeneity. To guide the empirical analysis, we propose a cognitive heterogeneity model.

6.1 Cognitive Heterogeneity Model

Heterogenity in behavior has been modeled by level-k reasoning models (see Stahl and Wilson, 1994, 1995; Nagel, 1995; Crawford and Iriberri, 2007) which focus on the interaction between agents whose depth of reasoning, as captured by an integer k, is heterogeneous.[14] In the basic version, a level-k thinker best responds to the assumption that all other agents are level-$(k-1)$ agents. The strategy used by level-0 agents is exogenously specified and the behavior of remaining agents is thus characterized recursively. Experimenters have found that for a variety of games (see Kawagoe and Takizawa (2012) for centipede games, see Crawford and Iriberri (2007) for auctions), given distributions of level-k types fit the data quite well. Level-k has also been used to describe behavior in cheap talk games (Cai and Wang, 2006; Wang et al., 2010; Kawagoe T, 2009) where the level-0 strategy of a sender (receiver) is assumed to be truth-telling (trusting).

Consider the following simple specification of the level-k model. Level-0 agents behave as in the naïve voter model: they truth-tell and vote for the decision indicated by the majority of announced signals. Level-1 agents best respond to the assumption that all others are level-0 agents and maximize individual payoffs. This involves lying after a contrary signal not only in minority but also in majority (whether in a heterogeneous or in a homogenous committee), as well as applying the type-specific payoff-maximizing decision rule. The lying incentive in minority echoes the profitable unilateral deviation scenario arising in the putative truthful-sincere equilibrium analyzed by Coughlan (2000). Lying bends the decision rule applied by majority types (if the lie is taken at face value) and therefore constitutes a profitable individual deviation. In majority and in a homogenous committee lying can bend the decision rule of level-0 types who wrongfully apply the majority heuristic instead of the optimal decision rule. This is only a profitable deviation if there is a high fraction of level-0 types. Lying in majority can be interpreted as benevolent, given that a lying subject improves the expected payoff of other subjects sharing his preference type.

The above model ignores important behavioral features that appear empirically relevant.

[14] See also Goeree and Holt (2004) for a related model of noisy introspection.

First, agents act noisily in response to beliefs, as captured, for example, by the popular quantal response model proposed in McKelvey and Palfrey (1995, 1998). Second, lying in majority is less intuitive than lying in minority. Finally, there is experimental evidence that subjects exhibit different levels of reasoning across different families of games (see Georganas et al., 2015). One way to rationalize this latter finding is to assume that the strategic sophistication level of an agent's behavior in a given game depends on the relation between the game's complexity and the agent's exogenous level of cognitive sophistication. An individual characterized by a given level of sophistication will exhibit a higher depth of reasoning, the simpler the game.

We incorporate some of the above concerns into a noisy version of the above introduced model of level-k thinking featuring level-1 agents that differ in their strategic sophistication. Let any level-1 agent exhibit a sophistication level s drawn from a distribution g with full support on $[0,1]$. Variable s determines the propensity of a level-1 agent to make errors. More precisely, let any s be associated with probabilities $l(z,s)$ and $d(s)$. The function $l(z,s)$ indicates the probability that a level-1 agent of sophistication level s lies after a contrary signal given that a total of $z \in \{1,2,3\}$ agents (him included) have his preference type in the committee. The function $d(s)$ indicates the probability that the level-1 agent applies decision rule $\Lambda(1)$ to the observed signal profile as opposed to $\Lambda(0)$, at the voting stage. We make the following extra assumptions. First, the above introduced functions are continuous and monotonically increasing in s, reflecting the fact that more sophisticated agents are less prone to make mistakes. Second, $l(z,1) > .5 > l(z,0), \forall z \in \{1,2,3\}$ and $d(1) > .5 > d(0)$, capturing the fact that a maximally (minimally) sophisticated level-1 agent is more (less) likely than not to act optimally whatever the committee composition. Third, $l(1,s) > l(2,s) > l(3,s), \forall s \in [0,1]$, reflecting the fact that lying is more intuitive the fewer agents share one's preference type. To close the model, we assume that level-1 agents always truth-tell after a conform signal. We assume that the committee only contains level-0 and -1 agents and therefore do not describe the behavior of higher order types. We summarize our prediction for the above introduced noisy level-k thinking model in the following proposition.

Proposition 4.

a. Private-Het and Private-Hom: Level-0 agents truth-tell and vote for the decision indicated by the majority of signals in the observed signal profile. Level-1 agents truth-tell after a conform signal. A level-1 agent of sophistication s applies $\Lambda\left(d(s)\right)$ *to the observed signal profile.*

b. Private-Het: After a contrary signal, (i) a minority level-1 agent of sophistication s lies with probability l(1,s), and (ii) a majority level-1 agent of sophistication s lies with probability l(2,s).

c. Private-Hom: After a contrary signal, a level-1 agent of sophistication s lies with probability l(3,s).

The above proposition implies a particular pattern of lying and voting rates in *private* treatments, as we explain below. If we classify subjects into categories on the basis of scenarios in which they lie consistently, a subject's category will be predictive of his sophistication level. We define four categories, C1-C4, for *Private-Het* and two categories, C5-C6, for *Private-Hom*. Consistent lying at a given information set is defined as lying more than 50% of the time. In the *Private-Het* category C1, agents lie consistently both in majority and minority. In C2 (C3) agents lie consistently only in minority (majority) and in C4 agents never lie consistently. In the *Private-Hom* category C5, agents lie consistently while in C6 agents do not.

Given the assumed behavior of level-k types, C3 behavior is not generated by the model and the group of subjects categorizable as C3 should thus be empty. To see this, recall that a level-1 agent of sophistication s is more likely to lie after a contrary signal if in minority than if in majority. The law of large numbers thus implies that if an agent consistently lies in majority, he must also consistently lie in minority. By the law of large numbers, categories C1, C2 and C5 should contain exclusively level-1 agents while categories C4 and C6 should concentrate all level-0 subjects as well as some level-1 agents. The proposition implies different average sophistication levels across categories C1, C2, C4 and C5. Let $E(s|Cx)$ denote the average sophistication level among Cx-agents. It must be true that

$$E(s|C5) > E(s|C1) > E(s|C2) > E(s|C4). \tag{1}$$

The intuition for the above is as follows. Let threshold s_r, for $r \in \{1, 2, 3\}$, correspond to the s-value at which $l(r, s)$ crosses the horizontal .5 line. Given that $l(1, s) > l(2, s) > l(3, s), \forall s \in [0, 1]$, it is trivially true that $s_3 > s_2 > s_1$. Now, simply note that C5 subjects are defined by $s \geq s_3$, C1 subjects are defined by $s \geq s_2$, C2 subjects are defined by $s_2 > s \geq s_1$ and C4 subjects are defined by $s \leq s_1$. In other words, the conditional distribution of s given C5 first order stochastically dominates (FOSD) the conditional distribution of s given C1. Similarly, the conditional distribution of s given C1 FOSD the conditional distribution of s given C2 and the same applies to C2 vs C4.

The above characterization implies a particular ranking of lying rates across categories. Let $E(l(1, s)|Cx)$ and $E(l(2, s)|Cx)$ denote the average lying rate in respectively minority and majority conditional on being a member of category Cx, for $x < 5$. Similarly, let $E(l(3, s)|Cx)$ denote the average lying rate conditional on being a member of category Cx, for $x \geq 5$. It must be true that $E(l(1, s)|C1) > E(l(1, s)|C2)$. This follows from using the fact that $l(1, s)$ is increasing in s together with the previously obtained inequalities defining the sophistication levels of members of the different categories. Recall that C1 subjects satisfy $s \geq s_2$ while instead C2 subjects satisfy $s_2 > s$.

Our characterization in contrast does not pin down the relative size of $E(l(3,s)|C5)$ and $E(l(2,s)|C1)$. Indeed, two effects oppose each other. On the one hand, C5 subjects are more sophisticated on average than C1 subjects as seen earlier (we call this the *selection effect*). On the other hand, for any given s it holds true that $l(2,s) > l(3,s)$, i.e. lying is more intuitive the fewer agents share one's preference type. When comparing a C5 and a C1 majority subject sharing the same sophistication level s, the C5 subject's probability of lying after a contrary signal is thus strictly lower than that of the C1 majority subject (we call this the *size effect*). Which of the two effects dominates is a priori unclear, so that we cannot make a clear prediction of the ordering of lying rates for C1 and C5 subjects.

Finally, the obtained characterization implies a particular ranking of voting rates across categories. Let $E(d(s)|Cx)$ denote the average rate of applying $\Lambda(1)$ (as opposed to $\Lambda(0)$) to the observed signal profile conditional on being a member of category Cx. It must be true that

$$E(d(s)|C5) > E(d(s)|C1) > E(d(s)|C2) > E(d(s)|C4). \tag{2}$$

This follows from using the fact that $d(s)$ is increasing in s together with the previously obtained inequalities concerning the s-levels of C5, C1, C2 and C4. More precisely, recall that C5 subjects are defined by $s \geq s_3$, C1 subjects are defined by $s \geq s_2$, C2 subjects are defined by $s_2 > s \geq s_1$ and C4 subjects are defined by $s \leq s_1$, where $s_3 > s_2 > s_1$. To see that $E(d(s)|C5) > E(d(s)|C1)$, note that we have $E(d(s)|C5) = E(d(s)|s \geq s_3)$ while $E(d(s)|C1) = \alpha E(d(s)|s \geq s_3) + (1-\alpha)E(d(s)|s_3 > s \geq s_2)$, for some $\alpha \in [0,1]$, where it is trivially true that $E(d(s)|s \geq s_3) > E(d(s)|s_3 > s \geq s_2)$.

On the basis of the above analysis, we derive the following hypotheses for the cognitive heterogeneity model.[15]

Set of Hypotheses 4.

a. The lying rate after a contrary signal in minority of C1 subjects is higher than that of C2 subjects.

b. The average frequency of a conform decision given one conform signal is highest for C5 subjects, followed by C1, C2, and C4 subjects. In other words, there is a significant correlation between consistently lying after conform signals and consistently applying $\Lambda(1)$ *to the observed signal profile.*

Last, we assume that there is a large majority of level-0 subjects among subjects, so that the assumption made by level-1 players is empirically approximately correct. This implies that their

[15] In addition to the predictions on treatment behavior, one can also derive predictions on the post-experimental tests: a) In the SCT, C1 subjects perform better than C2 subjects who themselves perform better than C4 subjects; b)C5 subjects have the highest IDT threshold, followed by C1, C2 and C4 subjects.

lying should be payoff improving.

6.2 Results: Disaggregating behavior

In what follows, we examine our experimental data in light of the predictions derived from the cognitive heterogeneity model. First, we explore whether there is indeed heterogeneity in behavior that is consistent with our dichotomy of level-0 and level-1 types. Subsequently, we examine whether the pattern of lying and voting rates replicates the pattern proposed in the set of hypotheses 4.[16]

6.2.1 Evidence of heterogeneous behavior

In what follows, we analyze (a) whether there is a fraction of subjects who consistently lie in *Private-Het* minority, *Private-Het* majority and *Private-Hom* after a contrary signal, (b) whether lying after a contrary signal is indeed payoff-increasing, and (c) whether a fraction of subjects consistently applies its type-specific optimal decision rule $\Lambda(1)$ to the observed signal profile in all treatments.

Table 6 depicts the share of consistent liars and non-liars. We define consistent lying at a given information set as lying at least 50% of the time. While after a contrary signal the share of subjects who never lied amounts to approximately two thirds, almost all subjects never lie after a conform signal (above 96% in all three scenarios). We find that 9.38% of *Private-Hom* subjects lie consistently after a contrary signal, 14,58% in *Private-Het* majority and 25% in *Private-Het* minority. The relative size of these three groups supports our conjecture that lying is more intuitive (and probable), the higher the number of subjects of the other preference type.

Table 6: Share of non- and consistent liars (in %)

	Private-Hom		Private-Het			
	contrary	conform	contrary		conform	
			min	maj	min	maj
Lied never	77.08	96.88	65.22	66.67	96.77	96.88
Lied consistently	9.38	0.00	25.00	14.58	1.08	0.00

In the cognitive heterogeneity model lying is payoff-increasing because only a small fraction of cognitively sophisticated subjects (level-1 subjects) seizes the available profitable lying opportunity. Table 7 reports results from mixed-effects regressions with profits per period as the

[16] We refer to Appendix A.3.4 and A.3.5 for the analysis of the predicted ordering in the post-experimental tests. We refer to these results when appropriate.

dependent variable. Regression (1) includes data from both *private* treatments (1) while (2) only includes data from *Private-Het*.

Regression (1) indicates that lying is generally profitable. On average, a lie increases payoffs significantly from 46.77 to 58.97 tokens. The non-significance of the *Hom* and *Lie Contrary*Hom* coefficients in (1) indicates that the profitability of lying does not depend on the group composition being homogeneous or heterogeneous. Moreover, the non-significance of the *Minority* and *Lie Contrary*Minority* coefficients in (2) indicates that the profitability of lying does not depend on being in majority or minority.[17]

With regards to the heterogeneity in voting behavior, we observe that across all treatments a significant proportion of subjects consistently applies $\Lambda(1)$ to the observed signal profile. In *Public-Het* and *Public-Hom* we find that a share of respectively 41.66% and 44.79% of subjects consistently (i.e. more than 50% of the time) votes for the conform decision given an observed signal profile containing a unique conform signal. In *Private-Het* and *Private-Hom*, these shares decrease to respectively 29.17% and 28.13%. The decrease in shares when going from *public* to *private* treatments naturally follows from the skepticism towards information retrieved through communication.[18]

6.2.2 Lying and voting behavior by categories

The following important caveat applies to this part of our analysis. In order to examine whether our data is consistent with the predicted order in lying and voting behavior, we simply compare empirical frequencies across selected categories of subjects and perform no statistical tests. One key reason is that we disaggregate behavior across subgroups of limited size, which implies low statistical power. Our exercise here is thus rather to be interpreted as a preliminary and exploratory analysis and not as statistical hypothesis testing.

Table 8 shows for all categories C1-C6 the following statistics: 1) the number of subjects in each category, 2) the lying rates after a contrary signal in minority and majority, and finally 3) the frequency of a vote for the conform jar given an observed signal profile containing one conform signal. In line with previous results, the categories C4 and C6 contain the vast majority of subjects, 72% in *Private-Het* and 91% in *Private-Hom*. According to the cognitive heterogeneity

[17] Individual lying implies a coordination problem. If two subjects of the same preference type and both holding a contrary signal lie simultaneously, the triggered shift in the decision rule will be excessive. In accordance with this observation, we indeed find a decrease in profits when there are two simultaneous lies. This scenario however only arise extreme rarely, i.e. in less than 2% of cases, this being a trivial consequence of the low aggregate lying rate and of the small committee size. See Appendix A.3.3 for an analysis.

[18] We find similar shares in the IDT where 27.34% of all subjects apply the optimal decision rule and 70.31% apply the majority heuristic. For further information see A.3.5.

Table 7: Lying and Payoffs

	(1) Private	(2) Private-Het
Lie Contrary	12.20***	12.70**
	(4.32)	(5.38)
Hom	4.52	
	(3.06)	
Lie Contrary*Hom	-4.71	
	(6.72)	
Minority		-0.02
		(3.56)
Lie Contrary*Minority		-2.08
		(8.47)
Constant	46.77***	46.83***
	(2.21)	(2.26)
Obs.	1,978	959
# of Groups	32	16
# of Ind.	192	96

Notes: This table reports coefficients using a linear panel model with mixed effects. Regression (1) includes a dummy *Lie Contrary* equal to 1 if lying after a contrary signal and 0 otherwise, a dummy *Hom* equal to 1 for *Private-Hom* and 0 for *Private-Het*, as well as an interaction term *Lie Contrary*Hom*. Regression (2) controls for being in minority (*minority*) and lying in minority (*Lie Contrary*Minority*). Standard errors in parentheses.*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

model, these correspond mostly to level-0 agents. C1 and C2 subjects together constitute roughly 25% of *Private-Het* subjects while C5 subjects constitute 9% of subjects in *Private-Hom*. These three categories of subjects correspond to level-1 agents in the model. Finally, the number of C3 subjects is very low (3%) as predicted.

In minority, the lying rate of C1 subjects (80.2%) is higher than that of C2 subjects (71.5%), as predicted. Recall that the intuition is that category C1 subjects exhibit a higher average level of sophistication than C2 subjects. In addition, we find that the lying rate of the C5 subjects is higher than that of C1 subjects (85.6% vs 77.6%), which suggests that the selection effect (i.e., being in C5 requires a higher sophistication level than being in C1) dominates the size

Table 8: Lying and voting behavior by categories

	Category	Obs	Lying rates		Vote
			Minority	Majority	Conform
Het	C1	11	80.2%	77.6%	66.9%
	C2	12	71.5%	10.4%	49.5%
	C3	3	6.7%	55.7%	42.6%
	C4	66	3.9%	3.5%	23.5%
Hom	C5	9		85.6%	70.3%
	C6	87		2.4%	25.1%

Notes: In *Private-Het*, we could not categorize 4 subjects because they did not receive a contrary signal in minority.

effect (i.e. lying as majority type in a heterogeneous committee is more intuitive than lying in a homogeneous committee). Finally, C4 agents have a very low average lying rate in both minority and majority, which is compatible with these being to a large extent level-0 agents who always truth-tell.[19] We also find the predicted ordering in the voting behavior: C5 vote more often conform than C1, C1 more often than C2 and C2 more often than C4.[20]

Result 4.

The lying rate after a contrary signal in minority of C1 subjects is higher than the one of C2 subjects. Similarly, C5 most often vote conform given an observed signal profile containing one conform signal, followed by C1, C2 and C4 subjects. There is thus a positive correlation between consistently lying after contrary signals and consistently applying $\Lambda(1)$ *to the observed signal profile.*

Summarizing, we find results that are consistent with the main predictions of the cognitive heterogeneity model. At the communication stage in *private* treatments, a small fraction of subjects (17% on average across treatments) consistently lies after contrary signals while the

[19] Appendix A.3.4 analyzes whether subjects in higher categories also perform better in the SCT. We find no clear ranking of performance in the SCT. Results suggest that subjects did not understand the (quite complex) SCT and therefore continued acting as in the treatment, although optimal behavior in the SCT deviates from optimal behavior in the treatment.

[20] We have repeated the analysis for two more demanding definitions of consistent lying, that place the bar at respectively 60% and 65%. Most results survive, in terms of Table 8. The only change is that the ranking of the rates of conform voting of C1 and C5 subjects is now reverted. These however remain close, and closer to each other than to remaining rates (in particular the C1 vs the C2 rates). Results are available on request.

vast majority of subjects always truth-tells. Across treatments, roughly 35% of subjects consistently apply their type-specific payoff-maximizing decision rule. Finally and most importantly, consistent lying after contrary signals is significantly associated with applying the type-specific payoff-maximizing decision rule. We thus identify two groups of agents whose behaviour corresponds approximately to that of respectively level-1 and -0 agents in the cognitive heterogeneity model.

7 Conclusion

This paper reports results from a 2x2 experimental design aimed at understanding the drivers of individual behavior in a simple communication and voting game featuring known heterogeneous preference types. Besides the standard model of self-interested and strategic agents, we also tested models of social preferences and of naïve voters. Aggregate behavior is not consistent with any of the models which share the feature that they assume homogenous agents. Further disaggregating results, we find heterogeneous individual behavior that is consistent with the predictions of a level-k model featuring two cognitive sophistication levels. The numerically dominant level-0 subjects truth-tell and predominantly vote with the majority of signals. In contrast, sophisticated subjects tend to apply their type-specific payoff-maximizing decision rule and lie in a way that allows them to influence the commitee's decision in their favor.

Our experimental findings caution against interpreting low lying rates as reflecting homogenous truthful communication (with a low uniform rate of lying). Rather, the lying rates are consistent with the presence of a small share of sophisticated consistent liars facing a large majority of unsophisticated truth-tellers. More broadly, this paper highlights the need to integrate cognitive heterogeneity into models of committee decision-making. The mechanism design literature has a growing body of works that assume deviations from rationality (e.g. no preference maximization (de Clippel, 2014), varying but bounded "depths of rationality" (Saran, 2016)). These may provide a starting point for future theoretical work on committee design.

Though this experiment finds no role for social preferences, richer deliberation processes may contradict this conclusion. Debate might in some cases stimulate empathy, solidarity and common identity while it may in other cases reinforce *in vs outgroup* dichotomies and cause preference polarization. Future experiments ought thus to examine other deliberation protocols (e.g. sequential, repeated, subgroup-based).

References

Austen-Smith, D. and J. S. Banks (1996). Information aggregation, rationality, and the condorcet jury theorem. *The American Political Science Review 90*(1), 34–45.

Austen-Smith, D. and T. J. Feddersen (2006). Deliberation, preference uncertainty, and voting rules. *The American Political Science Review 100*(2), 209–217.

Battaglini, M., R. B. Morton, and T. R. Palfrey (2008). Information aggregation and strategic abstention in large laboratory elections. *The American Economic Review 98*(2), 194–200.

Battaglini, M., R. B. Morton, and T. R. Palfrey (2010). The swing voter's curse in the laboratory. *The Review of Economic Studies 77*(1), 61–89.

Bock, O., I. Baetge, and A. Nicklisch (2014). hroot: Hamburg registration and organization online tool. *European Economic Review 71*, 117–120.

Bolton, G. E. and A. Ockenfels (2000, Mar). Erc: A theory of equity, reciprocity, and competition. *The American Economic Review 90*(1), 166–193.

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

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

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

Coughlan, P. J. (2000). In defense of unanimous jury verdicts: Mistrials, communication, and strategic voting. *The American Political Science Review 94*(02), 375–393.

Crawford, V. P. and N. Iriberri (2007). Level-k auctions: Can a nonequilibrium model of strategic thinking explain the winner's curse and overbidding in private-value auctions? *Econometrica 75*(6), 1721–1770.

de Clippel, G. (2014). Behavioral implementation. *The American Economic Review 104*(10), 2975–3002.

Deimen, I., F. Ketelaar, and M. T. Le Quement (2015). Consistency and communication in committees. *Journal of Economic Theory 160*, 24–35.

Devine, D. J., L. D. Clayton, B. B. Dunford, R. Seying, and J. Pryce (2001). Jury decision making: 45 years of empirical research on deliberating groups. *Psychology, public policy, and law 7*(3), 622.

Dickson, E. S., C. Hafer, and D. Landa (2008). Cognition and strategy: a deliberation experiment. *The Journal of Politics 70*(04), 974–989.

Dohmen, T., A. Falk, D. Huffman, U. Sunde, J. Schupp, and G. G. Wagner (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association 9*(3), 522–550.

Esponda, I. and E. Vespa (2014). Hypothetical thinking and information extraction in the laboratory. *American Economic Journal: Microeconomics 6*(4), 180–202.

Feddersen, T. and W. Pesendorfer (1998). Convicting the innocent: The inferiority of unanimous jury verdicts under strategic voting. *The American Political Science Review 92*(01), 23–35.

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

Fehr, E. and K. M. Schmidt (1999, Aug). A theory of fairness, competition and cooperation. *The Quarterly Journal of Economics 114*(3), 817–868.

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

Georganas, S., P. Healy, and R. Weber (2015). On the persistence of strategic sophistication. *Journal of Economic Theory 159*, pp.369–400.

Gerardi, D. (2000). Jury verdicts and preference diversity. *The American Political Science Review 94*(02), 395–406.

Gerardi, D. and L. Yariv (2007). Deliberative voting. *Journal of Economic Theory 134*(1), 317–338.

Gneezy, U., B. Rockenbach, and M. Serra-Garcia (2013). Measuring lying aversion. *Journal of Economic Behavior & Organization 93*(0), 293–300.

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

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

Grosser, J. and M. Seebauer (2016). The curse of uninformed voting: An experimental study. *Games and Economic Behavior 97*, 205–226.

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

Hafer, C. and D. Landa (2007). Deliberation as self-discovery and institutions for political speech. *Journal of Theoretical Politics 19*(3), 329–360.

Kawagoe, T. and H. Takizawa (2012). Level-k analysis of experimental centipede games. *Journal Of Economic Behavior & Organization 82*(2), 548–566.

Kawagoe T, T. H. (2009). Equilibrium refinement vs. level-k analysis: An experimental study of cheap-talk games with private information. *Games and Economic Behavior 66(1)*, 238–55.

Le Quement, M. T. (2013). Communication compatible voting rules. *Theory and decision 74*(4), 479–507.

Le Quement, M. T. and V. Yokeeswaran (2015). Subgroup deliberation and voting. *Social Choice and Welfare*, 1–32.

MacCoun, R. J. (1989). Experimental research on jury decision-making. *Science 244*(4908), 1046–1050.

Martinelli, C. (2006). Would rational voters acquire costly information? *Journal of Economic Theory 129*(1), 225–251.

McKelvey, R. D. and T. R. Palfrey (1995). Quantal response equilibria for normal form games. *Games and economic behavior 10*(1), 6–38.

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

Meirowitz, A. (2007). In defense of exclusionary deliberation: communication and voting with private beliefs and values. *Journal of Theoretical Politics 19*(3), 301–327.

Murphy, R. O., K. A. Ackermann, and M. Handgraaf (2011). Measuring social value orientation. *Judgment and Decision Making 6*(8), 771–781.

Nagel, R. (1995). Unraveling in guessing games: An experimental study. *The American Economic Review 85*(5), 1313–1326.

Pennington, N. and R. Hastie (1990). Practical implications of psychological research on juror and jury decision making. *Personality and Social Psychology Bulletin 16*(1), 90–105.

Persico, N. (2004). Committee design with endogenous information. *The Review of Economic Studies 71*(1), 165–191.

Rabin, M. (1993, Dec). Incorperating fairness into game theory and economics. *The American Economic Review 83*(5), 1281–1302.

Saran, R. (2016). Bounded depths of rationality and implementation with complete information. *Journal of Economic Theory 165*, 517–564.

Sommers, S. R. and P. C. Ellsworth (2003). How much do we really know about race and juries-a review of social science theory and research. *Chi.-Kent L. Rev. 78*, 997.

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

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

Van Weelden, R. (2008). Deliberation rules and voting. *The Quarterly Journal of Political Science 3*(1), 83–88.

Wang, J. T.-y., M. Spezio, and C. F. Camerer (2010). Pinocchio's pupil: using eyetracking and pupil dilation to understand truth telling and deception in sender-receiver games. *The American Economic Review 100*(3), 984–1007.

A Appendix

A.1 Additional tables

Table A1: Share of wrong committee outcomes in %

	Public-Hom		Public-Het		Private-Hom		Private-Het	
	Blue	Red	Maj Blue	Maj Red	Blue	Red	Maj Blue	Maj Red
Wrong outcome	21.96	29.65	24.10	22.75	24.38	28.15	30.25	24.64
True jar blue	8.40	39.54	20.05	28.94	19.13	38.95	27.24	25.17
True jar red	35.52	19.75	28.16	16.55	29.63	17.35	33.27	24.11

Notes: The row *Wrong Outcome* indicates the share of wrong outcomes over all decisions, the row *True jar blue (red)* depicts the share of wrong outcomes when the true jar is blue (red). Blue (red) indicates that the committees contained only blue (red) types. Maj Blue (red) indicates that the majority of types in the committee is blue (red).

A.2 Post-experimental tests

The following post-experimental tests were conducted: the strategic communication test (SCT), the individual decision test (IDT), the lying aversion test and the social value orientation (SVO) slider.

The SCT test evaluates subjects' ability to communicate strategically. It is only taken by subjects in the *private* treatments (as these involve communication) and quasi-replicates the treatment game. A subject keeps his preference type from the treatment. Other subjects are now substituted with computers whose known strategy is to truthfully announce their signals and vote sincerely under the assumption of truth-telling by others. In the SCT a subject only chooses his announcement in the communication stage. At the voting stage, he is replaced by a computer which votes sincerely on the basis of the subject's signal and others' (truthfully announced) signals. Payoffs obtained by the two computerized committee members are randomly allocated to two treatment participants. We use the strategy method to elicit all choices conditional on being in minority or majority and the available signal. In *Het* subjects face four scenarios: one is either in majority or in minority and one either holds a contrary or a conform signal. Of these, only the minority and contrary signal scenario provides a payoff-incentive to lie. In *Hom* subjects face two scenarios. The committee is homogeneous and one holds either a contrary or a conform signal. In both of these cases truth-telling is payoff-maximizing.

The second test is the IDT which evaluates the ability to choose the optimal decision rule.

A subject observes three signals as in the *public* treatments but now chooses a jar alone. As compared to the treatments, the IDT excludes effects related to beliefs about others' behavior or social preferences. We use the strategy method. Subjects make a decision for each of the four possible signal profiles, as we seek to identify the minimal number of conform signals required by a subject to choose the conform decision. A subject requiring a minimum of x conform signals to choose the conform decision is said to follow the threshold rule x. On the basis of IDT behavior, we assign threshold rule x to a given subject if the difference between 4 and his total number of conform decisions is x.[21]

The third test is a lying aversion test based on Gneezy et al. (2013). It is a two-player deception game where the sender's decision to lie increases own payment independent of the receiver's decision (see the original paper for more details). In contrast to Gneezy et al. (2013), any subject is assigned twice to a two-persons matching group and plays the game once as a sender and once as a receiver. We only use the decision made by subjects when acting as sender. We furthermore only let subjects play the game once in each role. Our test results replicate those of Gneezy et al. (2013).

The fourth test is a social value orientation slider aimed at measuring social preferences (Murphy et al., 2011). At the end of the experiment, subjects answered a questionnaire gathering information about their risk aversion, trust of others, and demographic characteristics. Subjects were also asked specific questions on how they played and underlying motives.

A.3 Additional analysis

A.3.1 Lying aversion

Lying behavior in the experiment may have been affected by agents' lying aversion, which was measured in the post-treatment lying aversion test. To test this hypothesis, we run three different regressions. Regression (1) is a discrete choice model with the dummy variable *lie given a contrary signal* as dependent variable. In regressions (2) and (3), we use a linear regression and take as dependent variable the number of lies during the 20 periods. In regression (2) we include all subjects, while in regression (3) we only include subjects who lied at least once. Regressions (1) and (2) allow us to test whether the independent variables influence respectively the probability to lie or the frequency of lying over the 20 rounds. In addition, we use the restriction *at least one lie* for regression (3), as we conjecture that subjects who lied at least once were more likely to identify the lying incentive. We use as independent variables the treatment dummy *Het*, *Period*

[21] Two caveats are in order. First, the assignment method rests on the assumption that subjects' decision rule is monotonic in the number of conform signals. Second, our method does not allow us to observe whether a subject's decision rule is stochastic as opposed to deterministic.

to control for learning effects, the dummy variable *SCT* to check for comprehension of lying incentives[22], *IDT treshold*, *lying aversion*, *SVO*, a dummy for subjects studying *Economics* and a gender dummy *Male*. We find that lying aversion exclusively influenced the behavior of those subjects who lied at least once. The variable *lying aversion* has no significant influence in either regression (1) or regression (2). As soon as we drop all non-lying subjects from regression (3), we find that lying aversion negative impacts the number of lies.

A.3.2 Risk attitude and decision-making

The post-experimental questionnaire contained a non-incentivized question on risk attitudes. This question was taken from the German SocioEconomic Panel (SOEP). Subjects were asked about their "willingness to take risks in general", and had to indicate their answer on a scale ranging from 0 ("risk averse") to 10 ("fully prepared to take risks"). This measure was found to highly correlate with incentivized measures on risk attitudes (Dohmen et al., 2011). The variable *risk* in the regression below corresponds to this measure.

To test whether risk attitudes influenced behavior, we run a regression for the *public* treatments where the dependent variable is a dummy equal 1 if the subject votes for the conform decision and 0 otherwise. Besides risk attitude (as retrieved from the post-experiment questionnaire), independent variables include subjects' IDT threshold, dummy variables for the number of conform signals, the A-levels math grade, the dummy variables *Economics* and *Male*. The coefficient for risk aversion is marginally significant ($p \leq 0.10$) and small in size, in contrast to the coefficients for the IDT threshold and the dummies for the number of conform messages. Note that the risk coefficient loses its significance once we introduce the gender dummy in regression (2). We find that both variables, *Male* and *Risk*, correlate significantly (correlation coefficient 0.19, $p < 0.01$). We therefore conclude that risk aversion had a negligible impact on behavior.

A.3.3 Potential Coordination Problem of Lying

As outlined in section 6.2.1, we report payoffs of (majority) types after respectively one and two simultaneous lies in Table A4 to analyze how payoffs depend on the number of simultaneous liars in a committee. In the *private* treatments we identify all aggregate signal realizations in which two subjects of the same preference type hold a contrary signal. These are the instances where two majority subjects would each have an incentive to lie unilaterally. We build matching group averages for profits after one lie and after two lies and compare profits. The table indicates that profits as expected decrease when shifting from one to two simultaneous lies. Crucially,

[22] We here use the answer from our first question in the SCT. Recall that it was rational to lie in *Private-Het*, but not in *Private-Hom*. The test is useful as a proxy for comprehension of lying incentives.

Table A2: Impact of lying aversion on lying behavior

	Lie	Number of lies	Number of lies
Het	1.24**	0.35	-0.99
	(0.51)	(0.31)	(0.61)
Period	0.05***		
	(0.02)		
SCT	3.94***	4.25***	3.38***
	(0.64)	(0.67)	(0.73)
IDT threshold	-0.85	-0.61	-1.27**
	(0.53)	(0.41)	(0.61)
Lying Aversion	0.04	0.03	0.12**
	(0.04)	(0.02)	(0.05)
SVO	-0.03	-0.02	-0.00
	(0.02)	(0.01)	(0.02)
Economics	1.01	0.60	0.45
	(0.64)	(0.59)	(1.04)
Male	0.72	0.43	-0.09
	(0.51)	(0.29)	(0.66)
Constant	-4.56***	1.42	4.43***
	(1.16)	(0.87)	(1.46)
Observations	1,978	192	65
# of groups	32	32	27
# of individuals	192	192	65

Notes: Regression (1) reports marginal effects calculated at the means of covariates using a logit panel model with mixed effects. The dependent variable is a dummy equal to 1 for a lie after a contrary signal and 0 otherwise. Regression (2) and (3) report coefficients using a linear regression model with standard errors clustered on matching group level. The dependent variable in regression (2) and (3) is the number of lies after a contrary signal over the course of the treatment by a given subject. In regression (3) we restrict the sample to subjects who lied at least once. Independent variables include a dummy *Het* equal to 1 for *Private-Het* and 0 for *Private-Hom*, *Period* taking the value of the corresponding period, the individual scores from the post-experimental tests, *SCT*, *IDT threshold*, *Lying Aversion* and *SVO*. *Economics* is a dummy variable taking the value of 1 for subjects studying economics, 0 otherwise. *Male* takes the value of 1 for male subjects, 0 for female subjects. Standard errors in parentheses.*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A3: Impact of risk attitude on decision-making in *Public*

	(1)	(2)
Het	0.19	0.18
	(0.26)	(0.26)
IDT threshold	-1.85***	-1.70***
	(0.26)	(0.26)
1 conform	2.30***	2.29***
	(0.19)	(0.19)
2 conform	7.62***	7.61***
	(0.33)	(0.33)
3 conform	9.40***	9.39***
	(0.64)	(0.64)
Risk	0.09*	0.07
	(0.06)	(0.06)
Math grade	-0.03	-0.03
	(0.13)	(0.13)
Period	0.05***	0.05***
	(0.01)	(0.01)
Economics		0.16
		(0.40)
Male		0.56**
		(0.28)
Constant	-0.76	-1.15*
	(0.66)	(0.68)
Observations	3,380	3,380
Number of groups	30	30

Notes: This table reports marginal effects calculated at the means of covariates using a logit panel model with mixed effects. The dependent variable is a dummy equal to 1 for a conform vote and 0 for a contrary vote. Independent variables include a dummy *Het* equal to 1 for *Public-Het* and 0 for *Public-Hom*, the *IDT threshold*, dummy variables for the number of conform signals, *risk*, the A-levels *math grade*, *period*, *Economics* (a dummy variable taking the value of 1 for subjects studying economics, 0 otherwise) and *Male* (taking the value of 1 for male subjects, 0 for female subjects).

however, two simultaneous lies only happen extremely rarely, i.e. in less than 2% of cases. For all practical purposes, a subject lying at the communication stage can thus legitimately assume to be the only one lying, as in the unilateral deviation scenario in the putative truthful-sincere equilibrium analyzed in Coughlan (2000).

Table A4: Profits of majority types per number of lies

		1 lie	2 lies
Private-Hom	Average profit	42.75	28.75
	Share of obs.	21.35%	1.69%
Private-Het	Average profit	35.13	20.00
	Share of obs.	29.14%	1.71%

Notes: Only lies by majority types after a contrary signal are counted in cases where two majority types hold a contrary signal.

A.3.4 Strategic Communication Test and Communication in the Treatments

After the *private* treatments subjects took the strategic communication test (SCT) aimed at checking their understanding of strategic lying. If, as argued, lying in the treatment was driven by superior cognitive ability, an intuitive conjecture would be that lying after contrary signals in the treatment correlates positively with better performance in the SCT.

Table A5: Lying in SCT in % by categories

			SCT lying			
	Category	#	only min	only maj	min & maj	never
Het	C1	11	9.09	18.18	72.73	0
	C2	12	16.67	25.00	8.33	50.00
	C3	3	33.33	66.67	0	0
	C4	66	0	12.12	6.06	81.82
Hom	C5	9	NA	77.78	NA	22.22
	C6	87	NA	5.75	NA	94.25

Recall that *Private-Het* and *Private-Hom* subjects did not take the exact same SCT. For *Private-Hom*, lying was never individually payoff-improving in the SCT. For *Private-Het* sub-

jects, the only scenario where lying was payoff-improving in the SCT was after a contrary signal in minority. Optimal communication behavior in the SCT thus differed from optimal behavior in the treatments.

Table A5 shows results in the SCT for each of the treatment groups C1-C6. We report lying rates conditional on contrary signals. No clear ranking of performance in the SCT emerges across the considered categories. The only striking regularity is that SCT behavior closely resembles treatment behavior for all categories but C2. For example, most C1 subjects (72.73%) lie both in minority and majority in the SCT, just as in the treatment. Similar insights apply to C3, C4, C5 and C6. Results suggest that subjects did not understand the (quite complex) SCT and simply continued acting as in the treatment (suggesting the presence of order effects), rendering SCT results little informative. In particular, the notion that other subjects were replaced by computers might have caused confusion.

A.3.5 Individual Decision Test and Voting in the Treatments

After the treatment subjects took the individual decision test (IDT) aimed at measuring their decision rule in an individual decision task. The idea is that the IDT gives a cleaner measurement for the decision rule than the treatment as it excludes the role of beliefs and strategic interactions. For the IDT comparison we pool subjects from all treatments since the proportion of individuals applying each IDT threshold does not differ significantly between treatments. In an ordered logistic regression featuring the IDT threshold as the dependent variable, the coefficients of all treatment dummies are insignificant ($p > 0.36$).

Table A6: IDT decisions by lying category

	Category	Obs	IDT
Het	C1	11	1.55
	C2	12	1.83
	C3	3	1.67
	C4	66	1.80
Hom	C5	9	1.11
	C6	87	1.80

Notes: In *Private-Het*, we could not categorize 4 subjects because they did not receive a contrary signal in minority.

IDT results reflect the heterogeneity in voting behavior, 0.78% of subjects have an IDT

threshold of 0, 27.34% of 1, 70.31% of 2 and 1.56% of 3, thus there are two large groups. More than two thirds of subject apply the majority heuristic and less than one third the optimal decision rule. The IDT threshold also correlates with treatment behavior. Subjects with an IDT threshold of 1 vote conform after one conform signal much more frequently than subjects with an IDT threshold of 2 (63% vs 24%). More importantly, in Table A6 we analyze the correlation of lying and IDT behavior. Based on their lying behavior subjects are classified into categories of a presumably higher sophistication which are presumably more likely to use the optimal decision rule in the IDT. As predicted, the IDT threshold characterizing the presumably very sophisticated C5 subjects is very low (1.1) and thus very close to the optimal threshold of 1. C5 subjects' threshold is lower than that of C1 *Private-Het* subjects (1.5). These in turn have a lower threshold than C2, C4 and C6 subjects (1.8). The only deviation from predictions by the cognitive heterogeneity model is the high threshold of C2 subjects (and thus suboptimal) in the light of their good performance in the treatment.

A.4 Instructions

We print instructions for the *Public-Hom* blue-biased type (B.1) and for the *Private-Hom* blue-biased type (B.2) treatments. Aspects where the instructions differ for red-biased types are indicated in round brackets. Aspects where the instructions differ for heterogeneous groups are indicated in square brackets. The subsequent parts mentioned in the instructions refer to the post-experimental tests. Instructions for post-experimental tests are available upon request.

A.5 Instructions Public-Hom blue-biased type

General explanations for the participants

You are taking part in an economic experiment. Please read the following instructions carefully. You can earn money in this experiment. Your payment will depend on your decisions and on the decisions of the other participants.

During the experiment communication is prohibited. Failure to comply will result in exclusion from the experiment and loss of earnings. Should you have any questions, please address them to us: hold your hand out of the cabin and one of the experimenters will come to your seat.

At the end of the experiment, all sums of money will be paid to you in cash. During the experiment monetary amounts do not correspond to Euro, but to points. In the end, the total point earnings that you obtained during the experiment will be converted into Euro, where: **150 points = 1 Euro.**

The study consists of four parts:

Part 1. Control Questions: you are asked to answer control questions to check comprehension.

Part 2. Experiment: The experiment consists of several parts. Your earning from all parts will be paid.

(1) The instructions for Part 1 can be found below.

(2) You will receive the instructions for the other parts later.

Part 3. End: After the experiment you will receive a questionnaire with general questions. Please fill this out carefully.

Part 4. Payment: You will receive the payment privately. The other participants will not know the amount of your payment.

Instructions Experiment Part 1

Part 1 of the experiment consists of 20 rounds. [At the beginning of the experiment, you will be randomly assigned to a type, type A or type B. The type allocation is maintained throughout the experiment.] In each round, all participants will be divided into groups of 3 participants randomly. [Per group there are either two Type A-participants and a Type B-participant or a Type A-participant and two type B-participants. You will be informed about the group composition at the beginning of each round.] The group allocation is renewed at the beginning of each round. Therefore the group composition changes in each round.

In the experiment you have the task to vote for one of two jars. There are two possible jars, which we call the Red and the Blue Jar. The Red Jar contains 7 red balls and 3 blue balls. The Blue Jar contains 7 blue balls and 3 red balls.

At the beginning of the game one of the two jars will be selected for your group at random. The probability that the Red Jar is selected is 50%. The probability that the

Blue Jar is selected is also 50%. You will not be told which Jar was selected. In Figure 1 you see the Red and the Blue Jar. Figure 2 displays the image of the unknown jar.

Figure 1: Red and Blue Jar

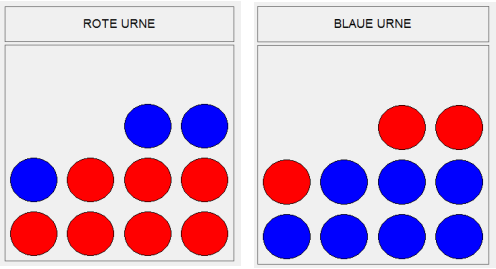
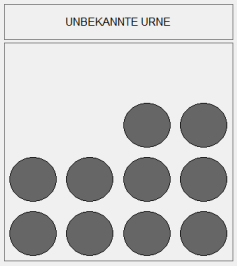


Figure 2: Unknown Jar



As information you will receive the color of three randomly drawn balls from the jar (see Figure 3). In three drawings one ball will be randomly drawn from the jar each time. Each drawing is carried out in two steps:

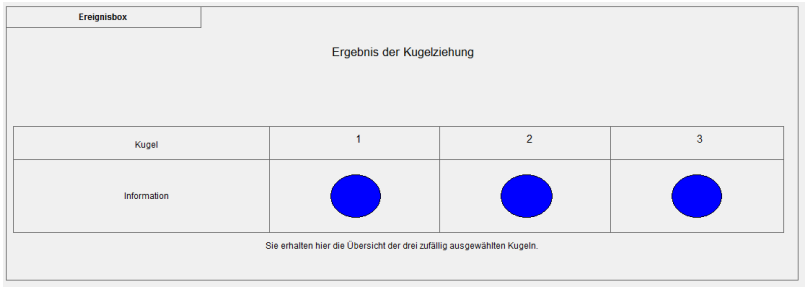
1. A ball is drawn from the jar.
2. The color is written down and the ball is immediately thrown back into the jar.

The number of balls in the jar thus remains the same at each draw. There are three drawings to obtain three balls. Each participant in your group receives the same three balls as information.

Differently colored balls may be drawn from the jar. However, all the balls are drawn from the same jar.

- When the **Red Jar** is selected for your group, each time a ball is drawn from a jar that contains 7 red balls and 3 blue balls.
- When the **Blue Jar** is selected for your group, each time a ball is drawn from a jar that contains 7 blue balls and 3 red balls.

Figure 3: Example for ball draw



After the ball draw the vote takes place. The vote is governed by the following rules:

- If the majority of participants votes for the Red Jar, your group decision is the Red Jar. If there are 2 to 3 votes for the Red Jar and 0 to 1 votes for the Blue Jar, the group decision is therefore the Red Jar.
- If the majority of participants votes for the Blue Jar, your group decision is the Blue Jar. If there are 0 to 1 votes for the Red Jar and 2 to 3 votes for the Blue Jar, the group decision is therefore the Blue Jar.

The payment you receive for the group decision depends on the accuracy of your group decision and of the actual jar.

- If your group decision **corresponds** to the selected Jar and the actual jar is the Red Jar, then you will receive 40 (160) points.
- If your group decision **corresponds** to the selected Jar and the actual jar is the Blue Jar, then you will receive 160 (40) points.
- If your group decision **does not correspond** to the selected jar, then you will receive 10 points.

Table 1: Payments

Number of votes for Red Jar	Number of votes for Blue Jar	Group Decision	Actual Jar	Payment [Type A]	[Payment Type B]
2 or 3	0 or 1	Red Jar	Red Jar	40 (160)	[160]
2 or 3	0 or 1	Red Jar	Blue Jar	10	[10]
0 or 1	2 or 3	Blue Jar	Red Jar	10	[10]
0 or 1	2 or 3	Blue Jar	Blue Jar	160 (40)	[40]

After all participants have voted, the votes will be counted and you will be informed about the outcome of the vote, i.e. votes for Red Jar, votes for Blue Jar, group decision, actual color of the jar and your payment. After the end of the round you will be assigned into new randomly selected groups and the next round begins.

You will receive the payments from all 20 rounds.

If you have questions about the experiment, please contact us now.

A.6 Instructions Private-Hom blue-biased type

General explanations for the participants

You are taking part in an economic experiment. Please read the following instructions carefully. You can earn money in this experiment. Your payment will depend on your decisions and on the decisions of the other participants.

During the experiment communication is prohibited. Failure to comply will result in exclusion from the experiment and loss of earnings. Should you have any questions, please address them to us: hold your hand out of the cabin and one of the experimenters will come to your seat.

At the end of the experiment, all sums of money will be paid to you in cash. During the experiment monetary amounts do not correspond to Euro, but to points. In the end, the total point earnings that you obtained during the experiment will be converted into Euro, where: **150 points = 1 Euro.**

The study consists of four parts:

Part 1. Control Questions: you are asked to answer control questions to check comprehension.

Part 2. Experiment: The experiment consists of several parts. Your earning from all parts will be paid.

(1) The instructions for Part 1 can be found below.

(2) You will receive the instructions for the other parts later.

Part 3. End: After the experiment you will receive a questionnaire with general questions. Please fill this out carefully.

Part 4. Payment: You will receive the payment privately. The other participants will not know the amount of your payment.

Instructions Experiment Part 1

Part 1 of the experiment consists of 20 rounds. [At the beginning of the experiment, you will be randomly assigned to a type, type A or type B. The type allocation is maintained throughout the experiment.] In each round, all participants will be divided into groups of 3 participants randomly. [Per group there are either two Type A-participants and a Type B-participant or a Type A-participant and two type B-participants. You will be informed about the group composition at the beginning of each round.] The group allocation is renewed at the beginning of each round. Therefore the group composition changes in each round.

In the experiment you have the task to vote for one of two jars. There are two possible jars, which we call the Red and the Blue Jar. The Red Jar contains 7 red balls and 3 blue balls. The Blue Jar contains 7 blue balls and 3 red balls.

At the beginning of the game one of the two jars will be selected for your group at random. The probability that the Red Jar is selected is 50%. The probability that the

Blue Jar is selected is also 50%. You will not be told which Jar was selected. In Figure 1 you see the Red and the Blue Jar. Figure 2 displays the image of the unknown jar.

Figure 1: Red and Blue Jar

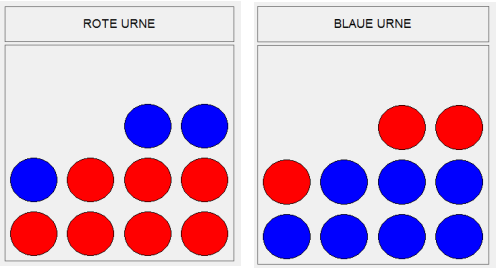
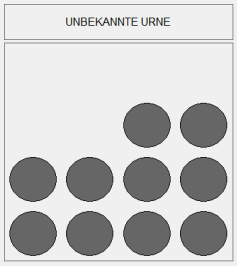


Figure 2: Unknown Jar



As information you will receive the color of three randomly drawn balls from the jar (see Figure 3). In three drawings one ball will be randomly drawn from the jar each time. Each drawing is carried out in two steps:

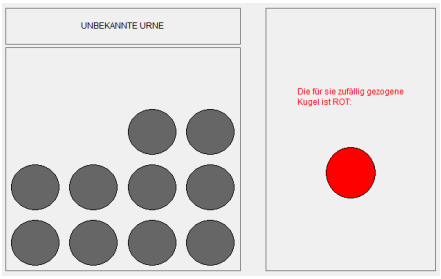
1. A ball is drawn from the jar.
2. The color is written down and the ball is immediately thrown back into the jar.

The number of balls in the jar thus remains the same at each draw.

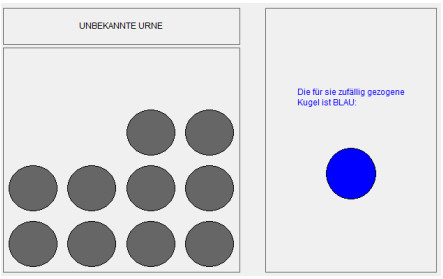
Figure 3 shows that two cases can occur. You either will be shown a red ball (case 1) or a blue ball (case 2).

Figure 3: Example for randomly drawn ball

Case 1) A red ball was drawn



Case 2) A blue ball was drawn

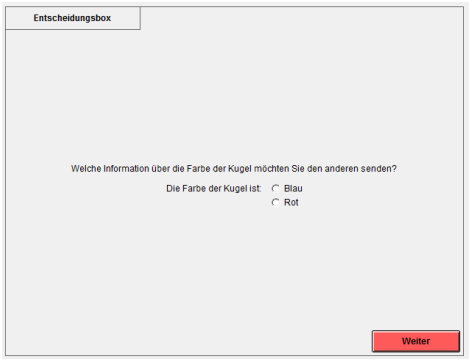


Differently colored balls may be drawn from the jar to the participants of the same group. However, all the balls are drawn from the same jar.

- When the **Red Jar** is selected for your group, each time a ball is drawn from a jar that contains 7 red balls and 3 blue balls.
- When the **Blue Jar** is selected for your group, each time a ball is drawn from a jar that contains 7 blue balls and 3 red balls.

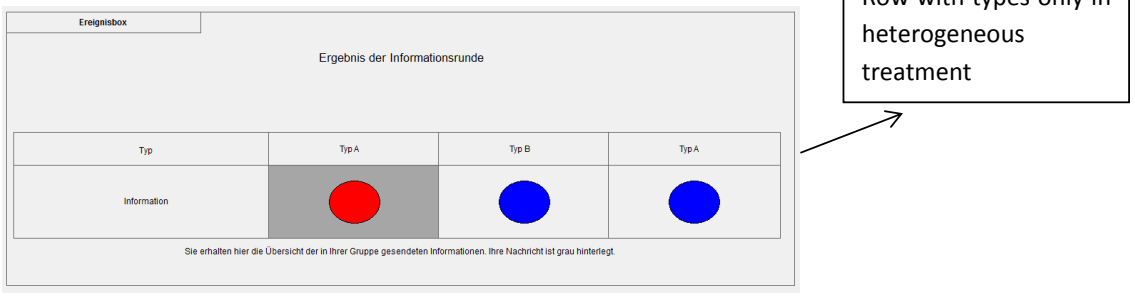
Now there is an information stage. You will send a message about the color of the ball that was shown to you to the other participants in your group. You can choose the content of the message independently of the actual color of the ball (see figure 4).

Figure 4: Message information stage



After you have sent the message, you receive the message of all the other participants of your group (see figure 5). In total, you see 3 messages, the messages of the other two participants and your own message.

Figure 5: Example for results of information stage



After the information stage the vote takes place. The vote is governed by the following rules:

- If the majority of participants votes for the Red Jar, your group decision is the Red Jar. If there are 2 to 3 votes for the Red Jar and 0 to 1 votes for the Blue Jar, the group decision is therefore the Red Jar.
- If the majority of participants votes for the Blue Jar, your group decision is the Blue Jar. If there are 0 to 1 votes for the Red Jar and 2 to 3 votes for the Blue Jar, the group decision is therefore the Blue Jar.

The payment you receive for the group decision depends on the accuracy of your group decision and of the actual jar.

- If your group decision **corresponds** to the selected Jar and the actual jar is the Red Jar, then you will receive 40 (160) points.

- If your group decision **corresponds** to the selected Jar and the actual jar is the Blue Jar, then you will receive 160 (40) points.
- If your group decision **does not correspond** to the selected jar, then you will receive 10 points.

Table 1: Payments

Number of votes for Red Jar	Number of votes for Blue Jar	Group Decision	Actual Jar	Payment [Type A]	[Payment Type B]
2 or 3	0 or 1	Red Jar	Red Jar	40 (160)	[160]
2 or 3	0 or 1	Red Jar	Blue Jar	10	[10]
0 or 1	2 or 3	Blue Jar	Red Jar	10	[10]
0 or 1	2 or 3	Blue Jar	Blue Jar	160 (40)	[40]

After all participants have voted, the votes will be counted and you will be informed about the outcome of the vote, i.e. votes for Red Jar, votes for Blue Jar, group decision, actual color of the jar and your payment. After the end of the round you will be assigned into new randomly selected groups and the next round begins.

You will receive the payments from all 20 rounds.

If you have questions about the experiment, please contact us now.