Le Quement, Mark T.; Marcin, Isabel

Working Paper

Communication and voting in heterogeneous committees: An experimental study

Preprints of the Max Planck Institute for Research on Collective Goods, No. 2016/5

Provided in Cooperation with:
Max Planck Institute for Research on Collective Goods

Preprints of the
Max Planck Institute for
Research on Collective Goods
Bonn 2016/5

Communication and voting in heterogeneous committees: An experimental study

Mark T. Le Quement
Isabel Marcin

MAX PLANCK SOCIETY

Preprints of the
Max Planck Institute
for Research on Collective Goods

Bonn 2016/5

Communication and voting in heterogeneous committees: An experimental study

Mark T. Le Quement / Isabel Marcin

March 2016

Max Planck Institute for Research on Collective Goods, Kurt-Schumacher-Str. 10, D-53113 Bonn
http://www.coll.mpg.de

Communication and voting in heterogeneous committees: An experimental study

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

March 22, 2016

Abstract

We study experimentally the effectiveness of communication in common value committees exhibiting publicly known heterogeneous biases. We test models assuming respectively self-interested and strategic-, joint payoff-maximizing- and cognitively heterogeneous agents. These predict varying degrees of strategic communication. We use a 2×2 design varying the information protocol (communication vs exogenous public signals) and the group composition (heterogeneous vs homogeneous). Results are only consistent with the third model. Roughly 80% of (heuristic) subjects truth-tell and vote with the majority of announced signals. Remaining (sophisticated) agents lie strategically and approximately apply their optimal decision rule.

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

JEL Classification: C92 · D72 · D82 · D83

[†]Institute of Microeconomics, University of Bonn, Adenauerallee 24-42, 53113 Bonn, Germany. E-mail: mlequem@uni-bonn.de

[‡]Max Planck Institute for Research on Collective Goods, Kurt-Schumacher-Str.10, 53113 Bonn, Germany, Email: marcin@coll.mpg.de

[§]We thank Christoph Engel, Guillaume Frechette, Sebastian Goerg, Jens Grosser, Oliver Kirchkamp, Sebastian Kube, Dimitri Landa, 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, NYU Abu Dhabi Behavioral Political Economy Workshop. 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 publicly known and 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). The problem at hand often has a common value dimension in the sense that members would agree on the right decision if the state of the world (whether a defendant is guilty or innocent, whether a reform will lower unemployment, whether a job candidate is competent) were known. Disagreement thus arises only if available information is not sufficiently clear-cut. If members exchange information before voting, full information sharing is desirable but standard game theory predicts that it might be unachievable because rational and self interested individuals have incentives to misrepresent their private information. We experimentally study the effectiveness of communication in a setup exhibiting the above-described features. While standard theory emphasizes the limits on information sharing, alternative theories stressing respectively the role of social preferences and cognitive constraints offer more positive prospects.

Consider a three-persons committee voting under simple majority rule. There are two possible underlying states, red and blue. Each agent privately receives an i.i.d. informative signal indicating either the blue or the red state. Members deliberate via simultaneous cheap talk (straw vote) in stage 1 before voting in stage 2. An agent's payoff depends on the group decision, the state of the world and his publicly known preference type (red-biased or blue-biased). The payoff-maximizing rule of a blue-biased (red-biased) agent is to vote for the blue (red) decision given at least one blue (red) signal in three. Instead, the total payoff-maximizing rule in a heterogeneous committee (whether containing two blue types and one red or vice versa) is to follow the majority of signals (majority heuristic).

If individuals simply maximize expected individual payoffs, there exists no equilibrium featuring truthful communication and sincere voting (so-called *TS equilibrium*) in which all agents truth-tell and vote sincerely (see Coughlan, 2000). The intuition is as follows. Assume two blue-biased and one red-biased agent. The decision rule applied in the TS equilibrium is the optimal decision rule of blue-biased agents, i.e. choose red only if three red signals are held. 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 and triggers the desired red decision. Truthful communication thus breaks down because agents act strategically and red and blue types are known to apply different decision rules. Experimental results (see Goeree and Yariv, 2011) however suggest that

this negative prediction might not be borne out in the laboratory. One can envisage two main explanations for this, which we now expose.

A first potential explanation is that subjects, though rational in the sense that they apply Bayes' rule and maximize their expected utility, have social preferences. If agents for example maximize the sum of members' individual payoffs, they all maximize the same objective function and the TS equilibrium accordingly exists. Clearly testing this theory would require observing how committees of varying compositions decide given three public signals. If the theory is correct, subjects in a homogenous committee should apply their type-specific payoff-maximizing rule while subjects in a heterogeneous committee should instead apply the average payoff-maximizing decision rule. Cognitive constraints are another potential reason for dominant truth-telling. Many agents might follow heuristics, be unable to communicate strategically or identify their payoff-maximizing decision rule. This would in turn leave a small fraction of sophisticated and self-interested subjects free to seize the generated strategic lying opportunities. Testing this theory would involve observing behavior in setups freed from strategic uncertainty and designed so as to evaluate in isolation individuals' performance in different tasks (strategic communication and optimal decision making).

We offer a first experimental investigation of the above game, which we formally term the *Condorcet Jury Model with publicly known heterogeneous preference types*. We test the three theories outlined above; the standard model of own payoff-maximizing strategic agents and two behavioral alternatives. The first alternative is a model of joint payoff maximization. The second is a model of k-level thinking featuring agents exhibiting varying depths of reasoning (level-0 and level-1). Level-0 agents are heuristic agents who truth-tell and vote with the majority of signals. Level-1 agents best respond to the assumption that all others are level-0 agents. We use a 2×2 experimental design which varies the committee composition as well as the information provision protocol. The first dimension determines whether the committee is heterogeneous or homogeneous (abbreviated respectively Het or Hom). The second dimension determines whether members all observe three public signals or instead each receive a private signal and engage in a round of simultaneous cheap-talk (abbreviated respectively Exo or Endo). Endo-Het is our main treatment while others are controls. We furthermore run individual post-experimental tests evaluating subjects' ability to communicate strategically and their individual choice behavior given public information.

We highlight the key predictions of our three models. We start with Model 1. As already mentioned, in the main Endo-Het treatment the TS equilibrium does not exist. In equilibrium, minority agents babble, majority agents truth-tell and all agents apply their type-specific payoff-maximizing decision rule when voting. Model 2 predicts that subjects truth-tell in both Endo

treatments and apply the joint payoff-maximizing decision rule in all treatments. As a consequence, the implemented decision rule changes when shifting from Endo-Het to Endo-Hom or from Exo-Het to Exo-Hom. Model 3's prediction for Endo-Het involves two types of behavior. Level-0 agents truth-tell and vote with the majority of publicly available signals. Level-1 subjects lie when holding a contrary signal (whether in minority or in majority) and apply their own payoff-maximizing decision rule conditional on available information.

We now describe our results. Model 1 is clearly rejected. We find no evidence of babbling by minority subjects in Endo-Het. In line with this finding, majority subjects condition their vote as much on minority subjects' announcements as on those of majority subjects. In Exo treatments, both preference types' average decision rule is heavily skewed towards the majority heuristic. Model 2 is also clearly contradicted. Though lying rates are as predicted low in all Endo treatments, applied decision rules differ from the average payoff-maximizing decision rules. In particular, we do not observe the predicted shift in decision rules between Exo-Het and Exo-Hom, which indicates that committee composition does not significantly affect subjects' objective function. Model 3 is instead roughly borne out by our findings once we disaggregate results further. At the communication stage in Endo treatments, an average of 17% of subjects consistently lies after contrary signals while the vast majority of subjects consistently truth-tells. Across treatments, roughly 35% of subjects apply their type-specific payoff-maximizing decision rule. Finally and most importantly, consistent lying after contrary signals is strongly associated with applying the type-specific payoff-maximizing decision rule. We thus identify two groups of agents who correspond roughly to level-1 and level-0 agents in Model 3.

Regarding the external relevance of our results, real committees are usually much larger than the committees studied here. The law of large numbers implies that larger committees will very likely contain some sophisticated agents despite a small probability that any given agent is sophisticated. Consequently, though we should expect to see dominant truth-telling in real committees, some degree of strategic communication is likely to always be present. Our results suggest that sophisticated agents are likely to have a strong impact on outcomes because their lies are taken at face value. Summarizing, strategic lying in committees appears to be an empirically relevant phenomenon that is likely to have significant welfare consequences.

Building on Condorcet's seminal essays on voting (see Condorcet, 1785), a formal 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 (2013); Battaglini et al.

(2008, 2010).

A set of newer contributions study the case of voting preceded by communication and have focused on the TS equilibrium. A milestone is the negative result obtained by Coughlan (2000) for the case of known heterogeneous preference types: If full truth-telling might lead to disagreement, then there exists no TS equilibrium. Le Quement and Yokeeswaran (2015) provides an equilibrium prediction for such committees. Assuming two preference types voting under unanimity rule, the authors find a unique responsive (symmetric and pure strategy) equilibrium given as follows. Jurors of the type endowed with veto power truthfully announce their signals, while remaining jurors babble and all agents vote sincerely. Deimen et al. (2015) offer a complementary analysis that assumes conditionally correlated signals. They find that for such information structures, the TS equilibrium is compatible with a positive probability of ex post disagreement. A parallel research agenda has been the extent to which uncertainty about preference types positively affects the possibility of communication (Austen-Smith and Feddersen, 2006; Meirowitz, 2007; Van Weelden, 2008; Thordal-Le Quement, 2013).

Voting with communication has also been examined experimentally.[1] Guarnaschelli et al. (2000) study a homogeneous jury and find, in contradiction with the intuitive prediction of full truth-telling, a small lying rate (around 5%) and skepticism towards information provided by others. Goeree and Yariv (2011)(GY in what follows) study the case of privately known (and potentially heterogeneous) preference types. The authors' primary objective is to test a theoretical prediction formulated in Gerardi and Yariv (2007), namely that all voting rules are equivalent given unrestricted communication. They find that subjects on average follow a simple heuristic which consists in truth-telling and subsequently voting with the majority of announced signals. Their theoretical prediction is furthermore verified. Our experiment complements GY by examining the case of publicly known heterogeneous preferences and by seeking to identify the drivers of individual behavior (e.g. social preferences, cognitive constraints). This involves introducing and testing alternatives to the standard strategic model. Given our simple communication protocol (straw votes) as well as the particularly explicit form of preference misalignment assumed (known preference types, two types), incentives for strategic communication should be relatively easy to identify for subjects (and arguably easier than in the GY setup). Furthermore, straw votes minimize the scope for the emergence of social preferences through communicative interaction. We therefore expect behavior to potentially be more strategic than in GY.

We proceed as follows. Section 2 presents our experimental design, introducing the four treat-

[1] 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.

ments as well as post-experimental tests. Section 3 introduces Models 1 and 2 with corresponding theoretical predictions. Section 4 analyzes aggregate behavior in the treatments with an eye to testing the predictions of Models 1 and 2. Section 5 introduces Model 3 and corresponding predictions. Section 6 disaggregates treatment behavior with an eye to testing the latter predictions. Section 7 concludes. The Appendix contains further discussions of our post-experimental tests as well as instructions.

2 Experimental Design

We here describe the treatments, the post-experimental tests and the experimental procedure.

2.1 The treatments

The basic setup is the same in all treatments and builds on the so-called Condorcet jury model. The state of the world takes one of two possible values, both being ex ante equally probable. The state is not observable. Each committee is composed of three subjects who choose between two alternatives by majority vote. In the jury interpretation, the group chooses between convicting and acquitting a defendant who is either guilty or innocent. We follow Guarnaschelli et al. (2000) and Goeree and Yariv (2011) in adopting a neutral description. The state is the (red or blue) jar selected by nature and the decision is either *red* or *blue*.

All treatments share the following basic timing. There is an information stage (stage 1) at which information regarding the jar color is received and possibly 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 jar selected by nature as well as their payoffs.

We vary the basic game on two dimensions across treatments, namely the information protocol (communication or exogenous information) and the preferences of subjects (heterogeneous or homogeneous). We consider two possibilities for each dimension and run one treatment for each entry of the thus obtained 2×2 matrix (see Table 1 below).

Table 1: Overview of treatments

		Preferences	
		homogeneous	heterogeneous
Information	endogenous	Endo-Hom	Endo-Het
	exogenous	Exo-Hom	Exo-Het

Though our primary objective is to explain behavior in the communication and heterogeneous preferences treatment, the three control treatments help us strengthen our test of competing the-

ories. Our fundamental underlying assumption here is that the same theoretical model explains subject behavior in all variations of the committee decision-making problem, and thus across all four treatments. If the model's prediction is borne out in the main treatment but contradicted by any of the control treatments, we shall accordingly reject it.

Information: In so-called *Exo* treatments, information comes in the form of three i.i.d. public signals. In *Endo* treatments, information is transmitted in two stages. In substage 1.a, each agent privately observes a signal. In the subsequent substage 1.b, each agent picks a simultaneously observed public message from the set $\{red, blue\}$. Messages are presented in a semi-anonymous way: Each is shown only with an indication of the subject's preference type. Note that a subject does not have the possibility to refrain from sending a message, as in the original Coughlan (2000) setup. 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. 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 $P(s = blue \mid red) = P(s = red \mid blue) = 1 - p = 0.3$. We shall repeatedly be referring to the *observed signal profile* of a subject in stage 1. In Endo treatments, it corresponds to a subject's own signal combined with the two signals announced by others. In Exo treatments, it corresponds to the three exogenous signals observed.

Preferences: Subjects' payoffs depend on the group decision and the realized jar. There are two possible preference types, *red* or *blue*, whose payoffs appear in Table 2. As discussed later, red (blue) types are biased towards the red (blue) jar. In so-called Hom (Het) treatments, preferences are homogeneous (heterogeneous). Jury composition is common knowledge at the start of the game. A subject whose preference type is (not) shared by some (any) other subject is called a majority (minority) subject. Given $j, -j \in \{red, blue\}$ and $-j \neq j$, we call a j-signal held by a type-j juror a *conform* signal and a $-j$-signal held by a type-j juror a *contrary* signal. We call decision j ($-j$) the *conform (contrary)* decision for a given type-j juror. If agents are risk neutral and self-interested, our payoff specification is equivalent to the one introduced in Feddersen and Pesendorfer (1998) and Coughlan (2000). In the latter 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. Note that payoffs are symmetric across red and blue types. Consider an outcome given by a profile of signals 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. It follows that we should expect identical behavior by red and blue types at symmetric information sets.

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	

We use a between-subjects design. At the start of the treatment, subjects are randomly assigned a preference type and a matching group of 6 subjects. 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 each period two committees are randomly formed, each with a different majority color in Het treatments. In Het treatments, each subject is thus very likely to experience multiple rounds in minority and in majority.

2.2 Post-experimental tests

At the end of the treatment session, subjects take a set of tests in the following order: (1) strategic communication test (SCT), (2) individual decision test (IDT), (3) lying aversion test and (4) social value orientation test. Payoffs from each of the tests are learned after the last test. While (3) and (4) are standard, (1) and (2) are introduced by us.

The SCT test evaluates subjects' ability to communicate strategically. It is only taken by Endo subjects and quasi-replicates the treatment game. A subject retains his treatment preference type. 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. An SCT subject only chooses his announcement. 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. An Endo-Het subject faces four scenarios. He is either in majority or in minority and either holds a contrary or a conform signal. Of these, only the minority and contrary signal scenario provides a payoff-incentive to lie.

An Endo-Hom subject faces two scenarios. The committee is homogeneous and he holds either a contrary or a conform signal. In both of these cases truth-telling is payoff-maximizing.

The second test is the individual decision test (IDT). A subject observes three signals as in Exo 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. 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. 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.

The third test is a lying aversion test based on Gneezy et al. (2013). We let every subject play the game once as sender and once as receiver. We refer to Appendix A.1 for a more detailed description. 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.

2.3 Experimental procedure

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. 96 subjects participated in each treatment, yielding 16 independent matchings groups per treatment. Subjects received written instructions which were read out loud by the experimenter (see Appendix B 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 before subjects were allowed to proceed to the actual game. 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.

3 Theoretical predictions

This section introduces two models and accompanying equilibrium predictions for the four treatments. Both models assume rational and risk-neutral agents while they differ on the assumed preferences. Agents maximize own payoffs in Model 1 and instead joint payoffs in Model 2. We focus on equilibria in symmetric strategies, in which agents with identical payoff functions use the same strategy. For each model, we derive from our theoretical predictions a set of testable conjectures. These contain point-predictions as well as directional hypotheses concerning treatment differences.

3.1 Model 1

This corresponds to the standard model analyzed in Feddersen and Pesendorfer (1998) and Coughlan (2000). We first recall the negative result of Coughlan (2000) for heterogeneous committees and subsequently present our equilibrium prediction for the treatments.

Given the payoffs in Table 2 and the assumption that agents only maximize own expected payoffs, an agent favors the conform decision 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 respectively 0,1,2 and 3 conform signals is given by respectively (approximately) $.07, .3, .7$ and $.93$. The optimal decision rule of each preference type given three signals is thus to choose the conform decision if at least one signal is conform. We denote by $\Lambda(x)$ the decision rule specifying the following probabilities of picking the conform decision after r conform signals in three: 0 if $r = 0$, x if $r = 1$ and 1 if $r \geq 2$. The rule $\Lambda(1)$ is thus the optimal decision rule of each type.

We start by recalling the impossibility result of Coughlan (2000) for our setting. If the committee contains at least one blue-biased and one red-biased agent, there exists no equilibrium in which all 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 Endo treatments, we focus on equilibria featuring maximal information sharing.

Proposition 1

a. Endo-Het: Majority (type) agents truth-tell while the minority (type) 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. Endo-Hom: All agents truth-tell. All agents apply $\Lambda(1)$ *to the observed signal profile.*

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

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

We add some intuition on our prediction for the Endo-Het treatment. 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 voting decisions. We now look at incentives at the communication stage. Here, the main intuition is that a majority agent recognizes that his optimal decision rule is implemented given the publicly pooled information. A detailed analysis of pivotality reveals that a majority agent's announcement is pivotal at a unique signal constellation and that the latter 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. The communication incentives of a minority agent are on the other hand trivial. Given that his announcement is ignored, he is indifferent between all messages and accordingly has no incentive to deviate from babbling. As to Endo-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 the following conjectures from Proposition 1.

Conjecture 1.1 *In Endo-Het, minority subjects babble and majority subjects truth-tell. In Endo-Hom, subjects truth-tell. Communication by minority Endo-Het subjects is less informative than communication by majority Endo-Het subjects and Endo-Hom subjects.*

In the above, we add a comparative statement because identifying intentional lying requires comparing lying rates to some base level lying rates by subjects who have no incentive to lie according to our equilibrium prediction. Subjects may be occasionally making mistakes in communication and thereby lie unintentionally. One needs to control for this base level noise before concluding that someone is intentionally being uninformative.

Conjecture 1.2 *In Endo-Het, majority subjects do not condition their vote on the announcement of the minority subject. In Endo-Het, majority subjects condition less their vote on the announcement of the minority subject than on that of a majority subject.*

We here again add a comparative statement. The rationale is that subjects might exhibit a base level skepticism towards others' messages because they anticipate that these make occasional mistakes in communicating. One needs to control for this effect before concluding that someone's messages are being ignored.

Conjecture 1.3 *In Exo-Het and Exo-Hom, subjects apply* $\Lambda(1)$ *to the observed signal profile. A given preference type applies the same decision rule in Exo-Het and Exo-Hom. In particular, the frequency of a conform vote given an observed signal profile containing one conform signal is the same in Exo-Hom and Exo-Het.*

We focus on Exo treatments because these by definition exclude the potential skepticism towards information arising as a consequence of communication. These treatments thus provide clean evidence of how subjects decide on the basis of information. We focus on behavior given a unique conform signal because we expect most of the variation in behavior (across subjects or treatments) to happen at this particular information set.

3.2 Model 2

We here assume that agents maximize the sum of committee members' individual type-specific payoffs, as specified in Table 2.[2] 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-biased agents and one (two) red-biased 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. We obtain the following equilibrium predictions. For Endo treatments we focus on equilibria featuring maximal information pooling, as in our analysis of Model 1.

Proposition 2

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

[2]The behavioral literature proposes different explanations for group induced preferences (e.g. social preferences, altruism, social norms) as well as different approaches to modeling these preferences. The literature on social preferences features outcome-based models that focus on inequity aversion or taste for efficiency, as well as intention-based models that highlight the role of reciprocity, kindness, etc. See for example Rabin (1993); Fehr and Schmidt (1999); Bolton and Ockenfels (2000); Charness and Rabin (2002).

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

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

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

Model 2 thus predicts truth-telling for any committee composition. Committee composition 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 (namely $\alpha \leq 0.699$ for minority Endo-Het subjects and $\alpha \leq 0.402$ for majority Endo-Het subjects). We derive the following conjectures from Proposition 2.

Conjecture 2.1 *All subjects truth-tell in both Endo-Het and Endo-Hom. Communication by majority Endo-Het subjects, minority Endo-Het subjects and Endo-Hom subjects is equally informative.*

Conjecture 2.2 *Exo-Het subjects apply* $\Lambda(0)$ *to the observed signal profile. Exo-Hom subjects apply* $\Lambda(1)$ *to the observed signal profile. Subjects apply different decision rules in Exo-Het and Exo-Hom. The frequency of a conform vote given an observed signal profile containing one conform signal is higher in Exo-Hom than in Exo-Het.*

In the above, recall that $\Lambda(0)$ and $\Lambda(1)$ only differ in terms of the behavior given a unique conform signal.

4 Aggregate behavior

In what follows, we empirically test the conjectures formulated in our theoretical predictions section. We start with conjectures derived from Model 1 and then proceed to those derived from Model 2.

We add a preliminary clarification regarding the possibility of pooling red and blue types in our analysis of communication and voting behavior. 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. Endo-Het majority, Endo-Het minority or Endo-Hom), a two-sided Mann-Whitney ranksum test (MW test in what follows) test does not reject ($p \leq .05$) the hypothesis that the frequency

of a lie is the same across the two preference types. Similarly, for each of the four treatments and each number of conform signals, a two-sided MW test does not reject ($p \leq .05$) the hypothesis that the frequency of a conform vote is the same across the two preference types. We accordingly systematically pool red and blue types in our data analysis.

4.1 Conjecture 1.1

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

Table 3: Lying rates in Endo treatments in %

Signal	Endo-Hom	Endo-Het
contrary	10.2	
conform	0.7	
contrary in min.		21.9
conform in min.		1.0
contrary in maj.		14.9
conform in maj.		0.5

We find no clear evidence of babbling by minority Endo-Het subjects. Given that these have a lying rate of approximately 0 after conform signals, babbling would imply that these roughly always lie after a contrary signal. To evaluate the conjecture that these babble, we test through a one-sided t-test whether their lying rate after a contrary signal is equal to 1. The test is rejected at a p-value of 0.0. For this and all following tests, we use matching groups as unit of independent observation.

We test the conjecture that minority Endo-Het subjects lie more after contrary signals than majority Endo-Het subjects. A one-sided Wilcoxon signed-rank test (WX test in what follows) marginally rejects ($p = 0.098$) the hypothesis that the lying rates of minority and majority subjects are equal. Likewise, we test the conjecture that minority Endo-Het subjects lie more after a contrary signal than Endo-Hom subjects. A one-sided MW test significantly rejects ($p = 0.02$) the hypothesis that their lying rates are equal. Concluding, we find weak evidence that minority Endo-Het subjects lie more than majority Endo-Het subjects. We find stronger evidence that minority Endo-Het subjects lie more than Endo-Hom subjects.

Result 1.1 *Minority Endo-Het subjects do not babble. Their lying rate after contrary signals is only slightly higher than that of majority Endo-Het and Endo-Hom subjects.*

We refer to Appendix A.1 for a short analysis of the impact of lying aversion (as measured in the post-experiment lying aversion test) on lying behavior in the treatments. In general, we find no significant impact of lying aversion on lying behavior.

4.2 Conjecture 1.2

Table 4 helps assess to which extent majority Endo-Het subjects condition their voting decision on the announcement of the minority Endo-Het subject. The table 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 types, one majority type and one minority type. For each possible signal held by the majority type, we examine four possible scenarios: Two involving a single conform message sent by others (either by the majority or the minority type), one involving no conform message sent by others, and one involving two conform messages sent by others. Choice frequencies show that the information provided by the minority type is influential (compare choice frequencies in cases 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 those of a majority type (compare cases 2 vs 3 and 6 vs 7).

Statistical analysis confirms that a minority Endo-Het subject's announcements are as influential as those of a majority Endo-Het subject. We test the conjecture that the frequency of a conform vote given the case 2 observed signal profile is larger than the frequency of a conform vote given the case 3 observed signal profile. A one-sided WX test does not reject ($p = 0.196$) the hypothesis that the frequencies of a conform decision are equal in cases 2 and 3. In the above, recall that Proposition 1.a) predicts a frequency of 1 in case 2 and a frequency of 0 in case 3.

We also test the conjecture that the frequency of a conform vote given the case 6 observed signal profile is equal to the frequency of a conform vote given the case 7 observed signal profile. 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. Recall that Proposition 1.a) predicts a frequency of 1 in cases 6 and 7. Here, our statistical analysis reveals that a minority type's announcement is actually even more influential than that of a majority type. A potential heuristic 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. Intuitively, the suspicious scenario is rather that of a minority type announcing a signal that confirms his bias.

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

Case	Own conform signal	Conform message by majority type	Conform message by minority type	Predicted frequency	Actual frequency	WS
1	0	0	0	0	4.0	
2	0	1	0	1	16.7	$p = 0.196^1$
3	0	0	1	0	12.4	
4	0	1	1	1	92.4	
5	1	0	0	1	44.4	
6	1	1	0	1	97.2	$p = 0.046^2$
7	1	0	1	1	100.0	
8	1	1	1	1	99.6	

Notes: WS is a Wilcoxon signed-rank test ([1]one-sided,[2]two-sided). The unit of independent observation is the matching group.

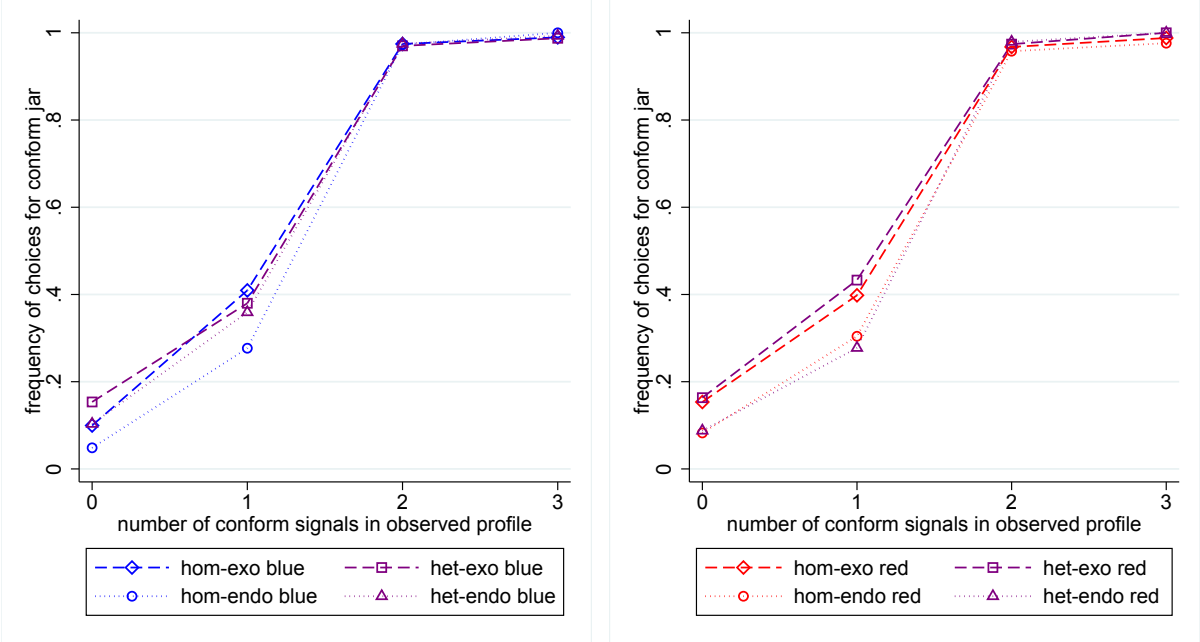
Result 1.2 *In Endo-Het, majority subjects do condition their vote on the announcement of the minority subject. Majority subjects condition their vote on the announcement of the minority subject more or less as much as on that of a majority subject.*

4.3 Conjecture 1.3

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 .35) given a unique conform signal. Average decision rules thus exhibit a reversal to the middle as compared to the $\Lambda(1)$. For each of the two Exo treatments, we separately test the conjecture that the average frequency of a conform vote given one conform signal is equal to 1. For both treatments, a one-sided t-test rejects ($p = .0$) the hypothesis that the frequency is equal to one.

As shown earlier, the average decision rules of red and blue types do not significantly differ, if one considers the probability of a conform vote as a function of the number of conform signals. This implies that if we represent the average decision rules of each type as the probability of a red vote conditional on the number of red signals in the observed profile of signals, the average decision rules of the two types unequivocally differ for each of the treatments. After one as well

Fig. 1: Frequencies of choices for the conform jar



(a) Blue types

(b) Red types

two red signals, the probability of a red vote by a red type is roughly .3 larger than that of a blue type.

We test the conjecture that the average frequency of a conform decision given one conform signal is the same in Exo-Hom as in Exo-Het. A two-sided MW test does not reject ($p = .89$) the hypothesis that the frequency is the same across the two treatments.

Result 1.3 *In Exo-Het and Exo-Hom, subjects do not apply* $\Lambda(1)$ *to the observed signal profile. Instead, both preference types apply roughly* $\Lambda(.4)$ *in both treatments. The average decision rule of subjects does not significantly differ across the two treatments.*

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 Exo 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 and the A-levels math grade. The coefficient for risk aversion is marginally significant ($p \leq .10$) 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 regression in Appendix A.2).

4.4 Conjecture 2.1

We here recall insights presented in our analysis of Conjecture 1.1. Table 3 shows that minority Endo-Het subjects, majority Endo-Het subjects and Endo-Hom subjects have a low lying rate after both conform and contrary signals. The lying rate after a conform signal is approximately 0 for all three types of subjects, while the lying rate after a contrary signal remains low in absolute terms. There is thus to a large extent truth-telling in Endo-Het and Endo-Hom. A one-sided MW test provides significant evidence that minority Endo-Het subjects lie more after contrary signals than majority Endo-Hom subjects. Heterogeneity thus causes a slight increase in lying.

Result 2.1 *Subjects to a large extent truth-tell in Endo-Het and Endo-Hom. But there is marginally more truth-telling in Endo-Hom than in Endo-Het.*

4.5 Conjecture 2.2

For the Exo-Het treatment, we test the conjecture that the average frequency of a conform vote given one conform signal is equal to 0. A one-sided t-test rejects ($p = .0$) the hypothesis that the frequency is equal to 0. In other words, Exo-Het subjects do not apply $\Lambda(0)$ to the observed signal profile. We also know from the analysis of Conjecture 1.3. that Exo-Hom subjects do not apply $\Lambda(1)$ to the observed signal profile. Finally, we know from the analysis of Conjecture 1.3. that average decision rules do not differ in Exo-Het and Exo-Hom and are given by roughly $\Lambda(.4)$.

Result 2.2 *The average decision rule does not significantly differ across Exo-Hom and Exo-Het. In both cases, it is roughly given by* $\Lambda(.4)$.

4.6 Summarizing insights

Model 1 is clearly rejected. We find no clear evidence of babbling by minority subjects in Endo-Het. Second, majority subjects condition their vote as much on minority subjects' announcements as on those of majority subjects. Third, in Exo treatments, subjects' average decision rule is not $\Lambda(1)$. Instead, the average decision rule is heavily skewed towards the majority heuristic. Model 2 is also clearly contradicted. We observe no significant shift in decision rules between Exo-Het and Exo-Hom, which indicates that committee composition does not significantly affect subjects' objective function.

5 A model of cognitive heterogeneity

While our two previous models assume fully rational agents, we now introduce a model of bounded rationality and heterogeneity in cognitive levels. The k-level thinking model Stahl and Wilson (see 1994, 1995); Nagel (see 1995); Crawford and Iriberri (see 2007) focuses on the interaction between agents whose depth of reasoning, as captured by an integer k, is heterogeneous. [3] In the standard version, a k-level thinker best responds to the assumption that all other agents are $(k-1)$-level 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 k-level types fit the data quite well.

We propose the following simple specification of the k-level model. Level-0 agents simply truth-tell and vote for the decision indicated by the majority of announced signals. These correspond to the subjects described in the experiment of Goeree and Yariv (2011). 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 whether in minority or majority and subsequently applying the type-specific payoff-maximizing decision rule. Note that the here described scenario echoes the profitable unilateral deviation scenario arising in the hypothetical truthful-sincere equilibrium analyzed by Coughlan (2000). Lying in majority could be described as benevolent lying as this improves the expected payoff of level-0 agents sharing the lying agent's preference type. We assume that the committee only contains level-0 and -1 agents and therefore do not describe the behavior of higher order types. We primarily base this assumption on the results of Goeree and Yariv (2011). We furthermore presume that level-1 agents constitute a small fraction of the total population.

The above model ignores important behavioral features. First, lying in majority is less intuitive than lying in minority. Second, agents act noisily in responding to beliefs, as captured for example by the popular Quantal-Response model proposed in McKelvey and Palfrey (1995, 1998). We propose a noisy version of the above introduced model of k-level thinking that accommodates both of these aspects. In examining our experimental data, we test the prediction of this noisy model rather than that derived from the non-noisy model.

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 $l(x,s), x \in \{1,2,3\}$ and $d(s)$. The function $l(x,s)$ indicates the probability that a level-1 agent of sophistication level s lies after a contrary

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

signal given that a total of x agents share 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 four above introduced functions are continuous and monotonically increasing in s, reflecting the fact that more sophisticated agents are less prone to make mistakes. Second, $l(x,1) > .5, \forall x \in \{1,2,3\}$ and $d(1) > .5$, capturing the fact that a maximally sophisticated level-1 agent is more 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 summarize our prediction for the above introduced noisy k-level thinking model.

Proposition 3

Level-0 agents truth-tell (if applicable) and vote for the decision indicated by the majority of signals in the observed signal profile. In Endo-Het and Endo-Hom, Level-1 agents truth-tell after a conform signal. A level-1 agent of sophistication s applies $\Lambda(d(s))$ to the observed signal profile.
a. Endo-Het: A minority (resp. majority) level-1 agent of sophistication s lies with probability l(1,s) (resp. l(2,s)) after a contrary signal.
b. Endo-Hom: A level-1 agent of sophistication s lies with probability l(3,s) after a contrary signal.

Proposition 3 implies a particular pattern of lying and voting rates in Endo treatments. Based on treatment behavior, one can classify Endo subjects into different categories depending on the scenarios in which they consistently lied. We define consistent lying at a given information set as lying more than 50% of the time. C1 agents lie consistently in majority and minority in Endo-Het. C2 (C3) agents lie consistently only in minority (majority) in Endo-Het. C4 agents never lie consistently in Endo-Het. C5 agents lie consistently in Endo-Hom while C6 agents do not.

By the law of large numbers, Proposition 3 implies that categories C1, C2 and C5 contain exclusively level-1 subjects while categories C4 and C6 concentrate all level-0 subjects as well as some level-1 subjects. Category C3 should 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.

Proposition 3 implies different average sophistication levels across categories C1, C2 and C5.

Let $E(s\,|Cx)$ denote the average sophistication level among Cx-agents. It must be true that

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

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 $s \geq s_3$ while C1 subjects are defined by $s \geq s_2$ and C2 subjects are defined by $s \in [s_1, s_2)$.

Double inequality (1) 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 directly from using (1) together with the fact that $l(1,s)$ is increasing in s, for $s \in [0,1]$. Double inequality (1) 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 (call this the selection effect). On the other hand, for any given s it holds true that $l(2,s) > l(3,s)$. When comparing a C5 and a C1 majority subject sharing the same s, the C5 subject's probability of lying after a contrary signal is thus strictly lower than that of the C1 majority subject (call this the size effect). Which of the two above effects dominates is a priori unclear.

Double inequality (1) also 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 (1) together with the fact that $d(s)$ is assumed increasing in s. We derive the following conjectures from Model 3. These concern treatment behavior as well as performance in post-experimental tests SCT and IDT.

Conjecture 3.1 *A fraction of minority Endo-Het, majority Endo-Het and Endo-Hom subjects lies consistently after a contrary signal.*

In the above, we define consistent lying at a given information set as lying more than 50% of the time.

Conjectures 3.2 *Lying after a contrary signal is payoff-increasing in Endo-Het majority, Endo-Het minority and in Endo-Hom.*

Recall that level-1 subjects assume that all others are level-0 subjects and conclude that lying after a contrary signal is payoff-improving. As already stated, we expect the proportion of unsophisticated agents (i.e. counterparts of level-0 agents) to be very large, which leads us to infer that lying after a contrary signal will indeed be payoff-improving in the treatments.

Conjecture 3.3 *The lying rate of C1 minority subjects after a contrary signal is higher than that of C2 minority subjects after a contrary signal.*

Conjecture 3.4 *In the strategic communication test (SCT), C1 subjects perform better than C2 subjects who themselves perform better than C4 subjects.*

Conjecture 3.5 *In all treatments and the IDT, a fraction of subjects consistently applies* $\Lambda(1)$ *to the observed signal profile. The share of such subjects in treatments is roughly the same as in the IDT.*

Before stating our last conjecture, we introduce new notation. Let f_x^T denote the average frequency of a conform decision given an observed signal profile containing a unique conform signal, the average being computed across subjects belonging to category Cx . Similarly, let t_x^{IDT} denote the average IDT-threshold among subjects belonging to category Cx.

Conjecture 3.6 *In Endo treatments,* $f_5^T > f_1^T > f_2^T > f_4^T$ *and* $t_5^{IDT} < t_1^{IDT} < t_2^{IDT} < t_4^{IDT}$. *In other words, there is a significant correlation between consistently lying after contrary signals in Endo-treatments and (1) consistently applying* $\Lambda(1)$ *to the observed signal profile in Endo treatments as well as (2) consistently applying a threshold of 1 in the IDT.*

6 Disaggregating behavior

In what follows, we test each of the conjectures derived from Model 3. A caveat is that our methodology in what follows mostly consists in comparing empirical frequencies across selected categories of subjects, as opposed to using statistical tests. One key reason is that we disaggregate behavior across subgroups of limited size, which implies low statistical power.

6.1 Conjecture 3.1

Table 5 is an expanded version of Table 3. The main new information is contained in the columns denoted ≥ 1, which indicate the average lying rate among subjects who lied at least once at a given information set. The idea is to seek evidence of the fact that subjects who lie once lie very frequently because subjects roughly divide into two categories, systematic truth-tellers

and systematic liars. Columns denoted *all* give the unconditional counterpart of the lying rate indicated in ≥ 1 columns. Columns denoted *share* indicate the share in % of subjects who lied at least once at a given information set.

For Endo-Hom, minority Endo-Het and majority Endo-Het subjects, the lying rate after a contrary signal increases strongly when going from the *all* column to the ≥ 1 column. The latter column features lying rates after contrary signals of at least 44.5%. The shares of subjects who lie at least once after a contrary signal in respectively Endo-Hom, Endo-Het minority and Endo-Het majority is given by 22.6%, 33.3% and 33.3%. In all three scenarios considered, the large majority of subjects thus never lies after a contrary signal while a small share of subjects instead appears to lie often.

We now directly identify subjects who lied consistently. We define consistent lying at a given information set as lying at least 50% of the time. We find that 8.3% of Endo-Hom subjects lie consistently after a contrary signal, 14.06% in Endo-Het majority and 24.5% in Endo-Het minority. The relative size of these three groups is consistent with our assumption that lying is more intuitive (and probable), the higher the number of subjects of the other preference type.

Table 5: Heterogeneous lying rates in Endo treatments in %

Signal	Endo-Hom			Endo-Het		
	all	≥ 1	share	all	≥ 1	share
contrary	10.2	44.5	22.9			
conform	0.7	22.3	3.1			
contrary in min.				21.9	63.0	33.3
conform in min.				1.0	29.8	3.1
contrary in maj.				14.9	44.7	33.3
conform in maj.				0.5	15.9	3.1

Notes: Column *all* indicates the overall lying rate, column ≥ 1 the lying rate of subjects that lie at least once in the corresponding category, and *share* indicates the share of ≥ 1 subjects for each category.

Result 3.1 *In Endo-Hom, Endo-Het minority and Endo-Het majority, a fraction of subjects lies consistently after a contrary signal while the vast majority of subjects always truth-tells after a contrary signal.*

6.2 Conjecture 3.2

In line with Model 3, we conjecture that the observed lying behavior reflects the fact that a small fraction of cognitively sophisticated subjects (level-1 subjects) seizes the available profitable lying opportunity. Table 6 reports results from mixed-effects regressions with profits as the dependent variable. Regression (1) includes data from both Endo treatments (1) while (2) only includes data from Endo-Het. 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 Endo-Hom and 0 for Endo-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*).

Our findings are as follows. 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.

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. We report payoffs of (majority) types after respectively one and two simultaneous lies in Table 7. We identify all Endo treatment 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, however, two simultaneous lies only happen extremely 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. 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).

Result 3.2 *Lying after a contrary signal is payoff increasing in Endo-Het majority, Endo-Het minority and Endo-Hom.*

6.3 Conjecture 3.3

Table 8 indicates treatment lying rates after a contrary signal for the categories C1-C6, showing furthermore the number of subjects in each category. In line with previous results, the categories

Table 6: Lying and Payoffs

	Endo	Endo-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. Standard errors in parentheses.*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 7: Profits of majority types per number of lies

		1 lie	2 lies
Endo-Hom	Average profit	42.75	28.75
	Share of obs.	21.35%	1.69%
Endo-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.

C4 and C6 contain the vast majority of subjects, respectively 72% in Endo-Het and 91% in Endo-Hom. According to Model 3, these correspond mostly to level-0 agents. C1 and C2 subjects together constitute roughly 25% of Endo-Het subjects while C5 subjects constitute 9% of subjects in Endo-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 by Model 3.

The lying rate in minority of C1 subjects (80.2%) is higher than that of C2 subjects in minority (71,5%), thus confirming Conjecture 3.3. Recall that the intuition is that categories C1 and C2 do not exhibit the same average level of sophistication. Although Proposition 3 does not predict any particular ordering of the lying rates of C5 subjects and C1 subjects in majority, we find that the lying rate of the former is higher than that of the latter (85.6% vs 77.6%). As previously discussed, the interpretation in terms of Model 3 is that C5 agents are so much more sophisticated than C1 agents that they lie more on average despite the fact that for any s, $l(2,s) > l(3,s)$. Using the vocabulary introduced after Model 3, the selection effect dominates the size effect. 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.

Table 8: Lying shares by Endo treatments lying category

	Category	Obs	Minority	Majority
Het	C1	11	80.2%	77.6%
	C2	12	71.5%	10.4%
	C3	3	6.7%	55.7%
	C4	66	3.9%	3.5%
Hom	C5	9	85.6%	
	C6	87	2.4%	

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

Result 3.3 *Minority C1 subjects have a higher lying rate after contrary signals than minority C2 subjects.*

6.4 Conjecture 3.4

After the Endo treatments, subjects took a strategic communication test (SCT) aimed at checking their understanding of strategic lying. We here examine the relation between treatment

lying rates and SCT results. Recall that Endo-Het and Endo-Hom subjects did not take the exact same SCT. For Endo-Hom, lying was never individually payoff-improving in the SCT. For Endo-Het subjects, 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. 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 9 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 (so-called order effect), rendering SCT results little informative. In particular, the notion that other subjects were replaced by computers might have caused confusion.

Table 9: Lying in SCT in % by Endo treatments lying category

		SCT lying			
Category	#	only min	only maj	min & maj	never
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
C5	9	NA	77.78	NA	22.22
C6	87	NA	5.75	NA	94.25

Conjecture 3.4 *C5 (C1)subjects do not unambiguously perform better in the SCT than C1 (C2) subjects. In the SCT, Endo subjects predominantly lie at those information sets where they consistently lie in treatments, suggesting the presence of an order effect rendering results of the SCT treatment uninformative.*

6.5 Conjecture 3.5

In Exo-Het and Exo-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 Endo-Het and Endo-Hom, these shares decrease to respectively 29.17% and 28.13%. Across all treatments, a significant proportion of subjects thus consistently applies $\Lambda(1)$ to the observed signal profile. The decrease in shares when going from Exo to Endo treatments can be explained by the skepticism towards information following from communication.

Next, we study individual decision test (IDT) results and reexamine treatment voting behavior as a function of IDT results. The first column of Table 10 presents the four IDT thresholds and the second column indicates the share of subjects belonging to each category, merging subjects across treatments.[4] Results echo those of the previous paragraph. Approximately 70 % of subjects share the same suboptimal IDT threshold of 2. The remaining 30 % mostly have a threshold of 1. For each IDT threshold, Table 10 reports the frequency of a conform vote for each possible observed signal profile in the treatments. The results reported in columns 3-6 show that behavior in the treatment parallels behavior in the IDT. 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%). We consider it unlikely that the IDT was subject to the same order effect as the SCT. The reason is that the IDT was very simple. Subjects were simply asked to pick a jar alone on the basis of three signals.

Result 3.5 *In all treatments and the IDT, a fraction of subjects consistently applies* $\Lambda(1)$*. The share of this category of subjects is roughly the same (*$\approx 35\%$ *) in the treatments and in the IDT.*

6.6 Conjecture 3.6

Table 11 shows treatment voting behavior as a function of treatment lying behavior, as described by the categories C1-C6. For category C1-C6, the column *Vote Conform* indicates the frequency of a vote for the conform jar given an observed signal profile containing one conform signal. We find the ordering predicted by Conjecture 3.6. The column *IDT* shows behavior in the individual decision test (IDT) as a function of treatment lying behavior. We roughly find the ordering predicted by Conjecture 3.6. The threshold characterizing 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

[4]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 10: Frequency of conform votes in % conditional on IDT threshold and observed signal profile

		# of conform signals in observed signal profile			
IDT	share in %	0	1	2	3
0	0.78	66.7	54.2	100	100
1	27.34	11.9	63.2	97.8	99.8
2	70.31	7.9	24.2	97.7	99.5
3	1.56	36.0	38.4	69.1	72.2

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

C1 Endo-Het subjects (1.5). These in turn have a lower threshold than C2, C4 and C6 subjects (1.8). The threshold of C2 subjects is surprisingly high (and thus suboptimal) in the light of their good performance in the treatment.

Result 3.6 *It holds true that* $f_5^T > f_1^T > f_2^T > f_4^T$. *Furthermore, it holds true that* $t_5^{IDT} < t_1^{IDT} < t_2^{IDT} \approx t_4^{IDT}$. *There is thus a significant correlation between consistently lying after contrary signals in Endo-treatments and (1) consistently applying* $\Lambda(1)$ *to the observed signal profile in Endo treatments as well as (2) consistently applying a threshold of 1 in the IDT.*

Table 11: Voting decisions by Endo treatments lying category

	Category	Obs	Vote Conform	IDT
Het	C1	11	66.9%	1.5
	C2	12	49.5%	1.8
	C3	3	42.6%	1.67
	C4	66	23.5%	1.80
Hom	C5	9	70.3%	1.1
	C6	87	25.1%	1.8

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

6.7 Summarizing insights

We find results that are consistent with the main predictions of Model 3. At the communication stage in Endo treatments, a small fraction of subjects (17% on average across treatments) consistently lies after contrary signals while the 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 strongly associated with applying the type-specific payoff-maximizing decision rule. We thus identify two groups of agents who correspond roughly to level-1 and -0 agents in Model 3.

7 Conclusion

We reported 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 cognitive heterogeneity. Aggregate behavior is neither consistent with the standard model nor with a model of joint payoff-maximization. Further disaggregating results, we find heterogeneous individual behavior consistent with two cognitive sophistication levels. The numerically dominant heuristic subjects truth-tell and vote in a way that is heavily biased towards the majority of signals. Sophisticated subjects instead lie in a way that allows them to favorably affect the implemented decision rule and subsequently approximately follow their type-specific payoff-maximizing decision rule.

Future experiments ought to examine other deliberation protocols (e.g. sequential, repeated, subgroup-based) and larger committees. 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.

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.

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

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

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.

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 (2013). The curse of uninformed voting: An experimental study. https://drive.google.com/file/d/0B1K0_OlGGtfPQ0dQTW1MTnVFSU0/view.

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.

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

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.

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.

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.

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

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

A Additional analysis

A.1 Lying aversion

The post-experiment lying aversion test 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).

Lying behavior in the jury experiment may have been affected by agents' lying aversion, which was measured in the post-treatment lying aversion test. To test this conjecture, 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. We in addition 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* to control for learning effects, the dummy variable *SCT* to check for comprehension of lying incentives[5], *IDT treshold*, *lying aversion* and *SVO*. 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.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.

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

Table 12: Impact of lying aversion on lying behavior

	Lie	Number of lies	Number of lies
Het	0.05**	0.30	-0.97*
	(0.02)	(0.30)	(0.56)
Period	0.001***		
	(0.000742)		
SCT	0.18***	4.48***	3.42***
	(0.02)	(0.70)	(0.77)
IDT threshold	-0.04*	-0.68	-1.28**
	(0.02)	(0.42)	(0.59)
Lying Aversion	-0.002	-0.03	-0.11**
	(0.002)	(0.02)	(0.04)
SVO	-0.001	-0.02	-0.01
	(0.001)	(0.01)	(0.02)
Constant		2.32**	6.71***
		(0.89)	(1.57)
Obs.	1,978	192	65
# of groups (cluster)	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. Standard errors in parentheses.*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 13: Impact of risk attitude on decision-making

	Exo
Het	0.17
	(0.26)
IDT threshold	-1.84***
	(0.26)
1 conform	2.24***
	(0.19)
2 conform	7.50***
	(0.32)
3 conform	9.30***
	(0.64)
Risk	0.10*
	(0.06)
Math Grade	-0.03
	(0.13)
Constant	-0.19
	(0.64)
Observations	3,380
Number of groups	30

Notes: This table reports marginal effects calculated at the means of covariates using a logit panel model with mixed effects.

B Instructions

We print instructions for the Exo-Hom blue-biased type (B.1) and for the Endo-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. Instructions for post-experimental tests are available upon request.

B.1 Instructions Exo-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

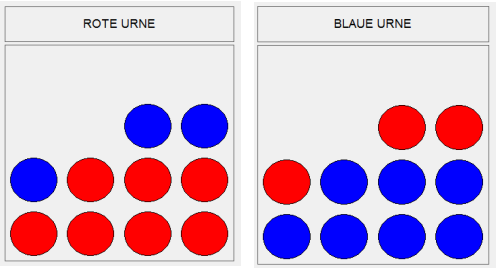
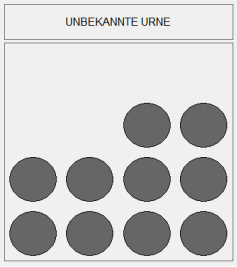


Figure 2: Unknown Jar



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:

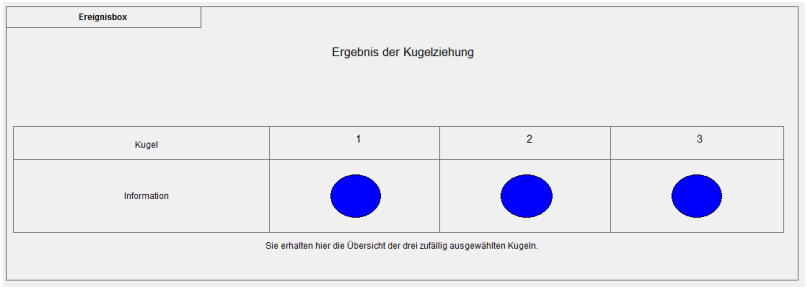
1. A ball is drawn from the jar.
2. The color is written down and the ball is immediately thrown back into the jar.

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.

B.2 Instructions Endo-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

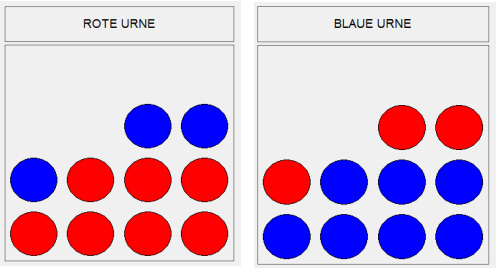
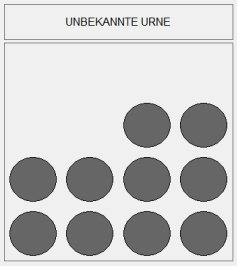


Figure 2: Unknown Jar



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:

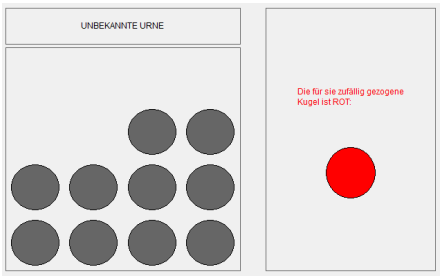
1. A ball is drawn from the jar.
2. The color is written down and the ball is immediately thrown back into the jar.

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

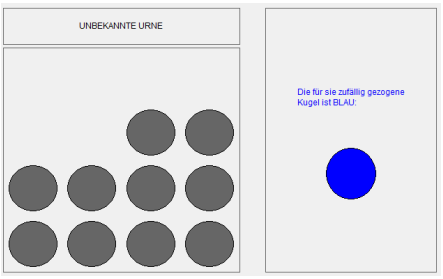
Figure 3 shows that two cases can occur. You either will be shown a red ball (case 1) or a blue ball (case 2).

Figure 3: Example for randomly drawn ball

Case 1) A red ball was drawn



Case 2) A blue ball was drawn

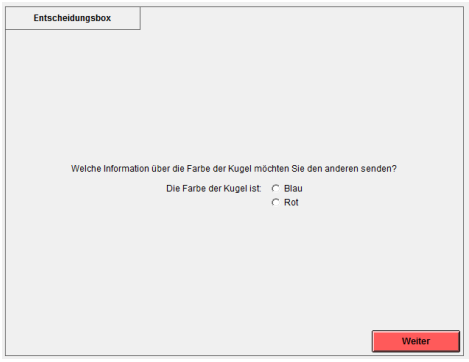


Differently colored balls may be drawn from the jar to the participants of the same group. However, all the balls are drawn from the same jar.

- 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.

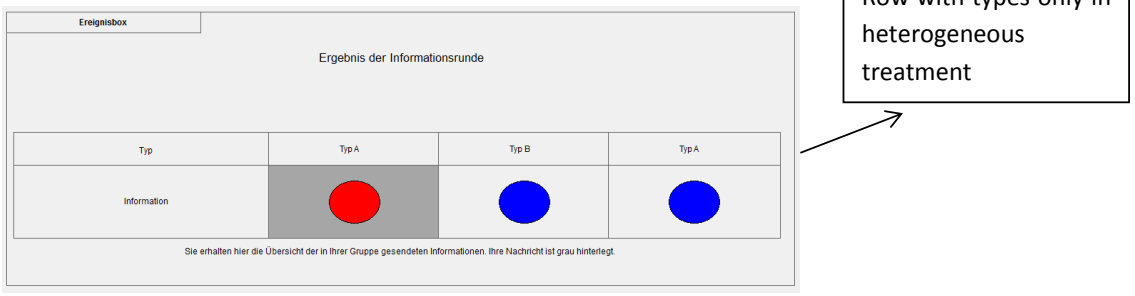
Now there is an information stage. You will send a message about the color of the ball that was shown to you to the other participants in your group. You can choose the content of the message independently of the actual color of the ball (see figure 4).

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.