January2018

Fondazione Eni Enrico Mattei

Working Paper

002.2018

The Strength of Weak Leaders - An Experiment on Social Influence and Social Learning in Teams

Berno Büchel, Stefan Klößner, Martin Lochmüller, Heiko Rauhut

Economic Theory
Series Editor: Matteo Manera

The Strength of Weak Leaders - An Experiment on Social Influence and Social Learning in Teams

By Berno Büchel, University of Fribourg, Economics
Stefan Klößner, Saarland University, Statistics and Econometrics
Martin Lochmüller, University of Hamburg, Economics
Heiko Rauhut, University of Zurich, Sociology

Summary

We investigate how the selection process of a leader affects team performance with respect to social learning. We use a lab experiment in which an incentivized guessing task is repeated in a star network with the leader at the center. Leader selection is either based on competence, on self-confidence, or made at random. Teams with random leaders do not underperform compared to competent leaders, and they even outperform teams whose leader is selected based on self-confidence. The reason is that random leaders are better able to use the knowledge within the team. We can show that it is the declaration of the selection procedure which makes non-random leaders overly influential. We set up a horse race between several rational and naïve models of social learning to investigate the micro-level mechanisms. We find that overconfidence and conservatism contribute to the fact that overly influential leaders mislead their team.

Keywords: Social Networks, Social Influence, Confidence, Overconfidence, Bayesian Updating, Naïve Learning, Sortition, Wisdom of Crowds

JEL Classification: D83, D85, C91

We thank Arun Advani, Sandro Ambuehl, Arun Chandrasekhar, Syngjoo Choi, P.J. Healy, Holger Herz, Matt Jackson, Michael Kosfeld, Friederike Mengel, Claudia Neri, and Muriel Niederle for helpful comments. Berno Buechel gratefully acknowledges the hospitality of the Economics Department of Stanford University and the financial support by the Fritz Thyssen Foundation. Heiko Rauhut acknowledges support by the SNSF Starting Grant BSSGI0 155981.

Address for correspondence:
Berno Büchel
University of Fribourg, Economics
Bd. de Pérolles, 90
1700 Fribourg
Switzerland
E-mail: berno.buechel@unifr.ch

The Strength of Weak Leaders – An Experiment on Social Influence and Social Learning in Teams*

Berno Büchel[1], Stefan Klößner[2], Martin Lochmüller[3], Heiko Rauhut[4]

[1] University of Fribourg, Economics, berno.buechel@unifr.ch

[2] Saarland University, Statistics and Econometrics, s.kloessner@mx.uni-saarland.de

[3] University of Hamburg, Economics, martin.lochmueller@gmail.com

[4] University of Zurich, Sociology, heiko.rauhut@uzh.ch

November 30, 2017

Abstract

We investigate how the selection process of a leader affects team performance with respect to social learning. We use a lab experiment in which an incentivized guessing task is repeated in a star network with the leader at the center. Leader selection is either based on competence, on self-confidence, or made at random. Teams with random leaders do not underperform compared to competent leaders, and they even outperform teams whose leader is selected based on self-confidence. The reason is that random leaders are better able to use the knowledge within the team. We can show that it is the declaration of the selection procedure which makes non-random leaders overly influential. We set up a horse race between several rational and naïve models of social learning to investigate the micro-level mechanisms. We find that overconfidence and conservatism contribute to the fact that overly influential leaders mislead their team.

JEL-Code: D83, D85, C91.

Keywords: Social Networks, Social Influence, Confidence, Overconfidence, Bayesian Updating, Naïve Learning, Sortition, Wisdom of Crowds

*We thank Arun Advani, Sandro Ambuehl, Arun Chandrasekhar, Syngjoo Choi, P.J. Healy, Holger Herz, Matt Jackson, Michael Kosfeld, Friederike Mengel, Claudia Neri, and Muriel Niederle for helpful comments. Berno Buechel gratefully acknowledges the hospitality of the Economics Department of Stanford University and the financial support by the Fritz Thyssen Foundation. Heiko Rauhut acknowledges support by the SNSF Starting Grant BSSGI0_155981.

1 Introduction

In our rapidly changing world, most modern organizations are embedded in highly dynamic environments. For the management of an organization, the first essential step to successful decision-making is the basic task of obtaining an accurate view of the environment.[1] For instance, this can be the foundation for defining a mission statement, as argued, e.g., in Bolton et al. (2013). Recently, there have been a number of contributions showing that organizations can improve their decision-making upon using the expertise of a single individual by harnessing the wisdom of crowds (e.g., Surowiecki 2004; Mannes 2009; Keuschnigg and Ganser 2017). However, this literature has not analyzed whether a team's ability to learn from each other depends on characteristics of the team leader.

Given the initial level of information of each team member, the accuracy of the updated opinions depends on the social learning process within the team. Many teams are organized such that one person, the team leader, directly communicates with each team member while the other members often communicate only indirectly with each other – via the team leader. In this paper, we address the question of *how the selection of the team leader affects the performance of social learning in the team.* Is it necessary that the central person is the one with the highest expertise? How does self-confidence affect the process of social learning? Should the selection criterion be declared or rather hidden? Answering these questions is important for the design of successful organizations.

To address these questions, we set up a lab experiment in which subjects are asked to answer incentivized estimation questions repeatedly. After each round, subjects can observe the guesses and the confidence levels of some of their team members according to a star network with the leader at the center. Thus, every team member observes the guesses of the leader, while only the leader observes the guesses of all members. We randomly allocate subjects into three treatments, which differ by the criterion that determines how the team leader is selected. In the baseline treatment (T0), the leader, i.e., the center, is selected at random. In the accuracy treatment (T1), the leader is the group member whose estimation of a related question was the most accurate in the team. Finally, in the confidence treatment (T2), the team member with the highest stated level of confidence (in the own answer of a related question) is selected. Potential ties in maximal accuracy or maximal confidence are broken at random.

Interestingly, a set of theoretical models following from the Bayesian approach to social learning predict for this setting that the selection of the center does not matter for the outcome, apart from the first two rounds, and that social learning is efficient.[2] The reason is that agents can exchange ("communicate") their opinions such that proper

[1] Indeed, disastrous decisions can often be traced back to management teams whose members are in disagreement, or – what is arguably even worse – who unintendedly agree on a distorted view of reality.

[2] For instance, Gale and Kariv (2003), Mueller-Frank (2013), and Rosenberg et al. (2009) provide frameworks to study social learning among rational agents who are Bayesian updaters.

aggregation leads to a common estimate (consensus) that is independent of who is at the center of the communication network. In contrast, a set of models of naïve social learning predict a strong impact of the center on the same process, which induces an inefficient outcome.[3] Based on the assumption that subjects fail to account correctly for the repetition of the center's and the others' initial opinion, they predict that consensus is approached over time, but with a strong "bias" towards the center's initial opinion. In particular, the center's weight on the consensus opinion is predicted to be proportional to her eigenvector centrality, which is several times larger than the other team members', in standard specifications of a naïve model of social learning. Unless the leader is much better infomed than the other team members, this is suboptimal, giving the leader's opinion too much weight. Now, any leader characteristic that further amplifies the weight of the leader's opinion undermines performance. As such we study the leader's self-confidence, as well as the declaration of why the leader was selected.

Results. In the experiment, we assess performance by the proximity of a guess to the correct answer. In particular, we measure the individual and the collective errors of the team's guesses, and use a measure of the wisdom of the crowds. Our first result is that leader selection based on accuracy (T1) does not outperform the random selection (T0), while leader selection based on confidence (T2) even undermines performance. The reason for this surprising result becomes apparent when isolating the effect of declaring how the leader is selected. The declaration of the leader as somewhat superior, be it in terms of past performance (T1) or of confidence (T2), induces the other team members to put more weight on the leader's opinion, making the team vulnerable to be misled by a single person. In contrast, teams with random leaders more equally weight each other's opinions with the consequence of a higher performance. On top of these effects, we assess how team performance is affected by (judgmental) overconfidence, which is the tendency to provide too narrow confidence intervals for one's estimates (e.g., Soll and Klayman (2004); Moore and Healy (2008); Herz et al. (2014)). It turns out that both overconfident leaders and overconfident other team members undermine performance, while overconfident leaders are worse. Hence, when designing a procedure for leader selection in a situation in which social learning is important, declared random selection is a viable option and overconfidence should be avoided.

In the second part of the paper, we set up a horse race between different models of social learning to shed more light on individual learning behavior and on the mechanisms of how leader selection affects the wisdom of crowds in networks. Despite a long tradition of theoretical insights and a growing body of empirical research, social learning behavior is still far from being fully understood. In line with the previous literature, we observe

[3]For instance, DeGroot (1974), Friedkin and Johnsen (1990), DeMarzo et al. (2003), Golub and Jackson (2010), and Acemoglu et al. (2010) study social learning among naïve agents.

that simple models (of naïve social learning) generally fit better than sophisticated models (of Bayesian social learning). This has the consequence that the leader's weight on the long-term opinion is large already due to her central position in the network structure. Moreover, the experimental data reveal that an important pattern is missing in both theoretical approaches: People tend to adapt their opinion less than predicted, a pattern called *conservatism*. Conservatism is a very common finding in experiments on belief updating and can be caused by (judgmental) overconfidence, as we show in this paper.[4] We incorporate this feature, which is missing in the theoretical literature on social learning, into both model classes and observe that incorporating conservatism improves the model fit of both model classes. Hence, there is important feedback from our data to theory development. Incorporating conservatism is not only a behavioral twist that matches empirical findings, but it also gives an additional reason for why overconfident leaders undermine performance.

Methodological Approach. In laboratory experiments (and in lab in the field experiments), theoretical models can be directly tested. For instance, Corazzini et al. (2012), Grimm and Mengel (2016), and Chandrasekhar et al. (2016) compare sophisticated models of Bayesian learning with simple models of naïve learning in settings in which their predictions diverge. The common conclusion is that the observations are more often consistent with the simple models. Similarly, Choi et al. (2005), Çelen and Kariv (2005), and Çelen et al. (2010) study predictions of social learning models experimentally. A caveat of this theory-testing approach is that the participants are confronted with highly stylized tasks such as guessing an average (or its sign) of randomly drawn numbers (Corazzini et al., 2012; Çelen and Kariv, 2005; Çelen et al., 2010) or finding an abstract true state (Choi et al., 2005; Grimm and Mengel, 2016; Chandrasekhar et al., 2016). It is questionable how the investigated learning behavior transfers to settings with real questions. A lab in the field approach (Chandrasekhar et al., 2016) does not fully mitigate this issue of external validity, because the types of questions are still often stylized. At the other side of the spectrum, real teams could be studied in the field to address our main research question (without the issue of external validity). However, besides the issues of noise, missing values, and the problem to measure performance of social learning, there would be a severe endogeneity problem. First, because face-to-face interaction gives rise to effects (e.g., due to charisma), which are difficult to control for, and second, because there is usually no proper randomization on who becomes a leader. For these reasons, we decided to use some middle ground between these two approaches (theory-testing experiment and

[4]Experiments on belief updating frequently find that real people are more conservative updaters than the theoretical model would predict (Mobius et al., 2011; Ambuehl and Li, 2014; Mannes and Moore, 2013), a pattern that has already been summarized in a classic survey (Peterson and Beach, 1967): "when statistical man and subjects start with the same prior probabilities for two population proportions, subjects revise their probabilities in the same direction but not as much as statistical man does[.]"

field data) in order to complement them. For that purpose, we imported a method developed outside of economics which has been increasingly used recently (Lorenz et al., 2011; Rauhut and Lorenz, 2011). Participants are asked to answer knowledge questions about vaguely known facts for which the true answer is known (and could in principle easily be looked up, e.g., on Wikipedia.com). The questions cover various topics and create a natural uncertainty among the participants who are paid according to their answers' accuracy. Arguably, teams who are able to estimate such factual questions accurately are also better at estimating states that cannot be simply measured, or at estimating future states of the world (which is of high relevance in real managerial or political teams). In our experiment, however, the quality of social learning can be assessed without waiting for the future to realize. The realism of that approach already changes the way subjects communicate with each other because, given that there is no stylized draw of signals which is common knowledge, it becomes important not only to communicate the guess, but also the own confidence in the guess. We consider it as a realistic assumption that people can "tag" the pieces of information they pass on with a confidence level by stating how confident they feel about their own guess.[5] This aspect is missing in most other experiments of social learning because it is simply not necessary to communicate confidence if signal quality is artificially made common knowledge.

Contribution. Our paper entails three contributions. First, we provide empirical evidence for the superiority of a selection procedure that is based on random leader selection ("sortition"). For both corporate and political governance, *sortition* (also called demarchy, allotment, or aleatory democracy) is discussed as an alternative selection procedure, which has its roots in ancient Athens and medieval Italy (Zeitoun et al., 2014; Frey and Osterloh, 2016). Despite a long list of claimed advantages of this procedure, empirical evidence showing its superiority is very rare. One exception is the study by Haslam et al. (1998), which shows experimentally that randomly selected leaders can enhance team performance in a task of deciding upon priorities in a hypothetical survival situation (e.g., after a plane crash). The mechanism behind the effect, however, remains largely unclear.[6] Our results not only show that random selection can be beneficial compared to selection based on confidence, but also demonstrate that it is the declaration of randomness rather than the selection at random per se that is the crucial aspect. The strength of random selection is based on the fact that the leader's influence on team members is not amplified by declaring the leader's specialty. Since the leader is already special because of her network position, additionally highlighting the leader's properties

[5]This is similar to the literature that considers "tagging" pieces of information with their source (Acemoglu et al., 2014; Phan et al., 2015).

[6]Interestingly, they also observe that randomly selected leaders are, despite their superior performance, often perceived by their team members as less effective than formally selected leaders.

by declaring them to be relevant for selecting the leader makes the others disrespect their own opinions, which results in a loss, because the wisdom of crowds is not harnessed.

Second, our experimental data reveal that the extent of (judgmental) overconfidence, i.e., providing too narrow confidence intervals, has a strong deteriorating effect on team performance. Indeed, overconfident team members undermine performance, and overconfident team leaders have an even stronger deteriorating effect. For the selection of leaders within organizations, this suggests that either overconfident leaders should be generally avoided or that there is at least a trade-off between beneficial effects of a leader's overconfidence (e.g., to foster coordination, Bolton et al. 2013, or to motivate team members, Gervais and Goldstein 2007) and the negative effect on social learning. (Judgmental) overconfidence may be partially domain-specific and state-dependent, but to some extent it is a personality trait that can easily be assessed, e.g., in an assessment center in the course of a selection procedure.

Third and finally, our paper makes a methodological contribution. By combining the experiments on factual questions with the theories on social learning, we bridge between neat theoretical frameworks and experimental set-ups that are less stylized (than those used for pure theory testing). By building this bridge it becomes apparent that the assumption of common knowledge about signal precision is problematic. Arguably, in reality people do not know the signal precision of their interaction partners, but form expectations about it, given what they know about this person and given how this person "tagged" her piece of information with a level of confidence. (Judgmental) overconfidence, as well as mistrust or anchoring effects, can lead to *conservatism* in updating, i.e., agents incorporate new pieces of information less than theoretically predicted. Bolton et al. (2013) further argue that other behavioral biases such as a selection bias in information acquisition can also induce conservatism (what they call resoluteness). We incorporate this idea into both naïve and rational models of social learning and find that the model fit of each model increases. This is informative for economic theory on naïve and rational social learning by opening a fruitful avenue for an empirically important model extension. In particular, our simple extensions of the models alter the prediction that consensus is reached or approached. Instead, they predict a persisting diversity of opinions, in which each agent's long-run opinion is "biased" in the direction of his initial opinion. This qualitatively different prediction could be studied more generally and be tested in follow-up experiments.

2 Experimental Design

In a nutshell, participants in this experiment were asked to answer the same knowledge questions multiple times in a row. The team leader could observe the previous answers of all team members, while the team members could only observe the previous answer of

the team leader. Treatments differed by the selection criterion that determined the team leader.

The experiment was conducted at the University of Hamburg and consisted of eleven sessions with a total of 176 subjects.[7] In each session, participants were randomly allocated into groups of four. The basic task was to answer a factual question individually and to provide a level of confidence for the answer. The closer the estimate was to the correct answer, the more it was honored by game points which were translated into actual payouts, as detailed in Table C.4. On average, sessions lasted for one hour and participants earned 9.50 Euros, which was close to the norm of the lab. The maximum feasible payout was 48.20, while the minimum was the show-up fee of 5 Euros. This fact was explicitly stated to the participants in order to highlight that the payout strongly depended on individual performance. It was pointed out verbally and in the written instructions that the use of mobile phones, smart phones, tablets, or similar devices would lead to expulsion from the experiment and exclusion from all payments.

Each session consisted of two phases: A selection phase I and a decision phase II. In the selection phase I, each participant answered eight different factual questions once. At the end of the experiment, one of these questions was randomly selected to be payoff-relevant. In the decision phase II, there was another set of eight questions, each of which was similar to one of the questions of the selection phase. For instance, there was a question about voter turnout in both phases of the experiment. Similarly, there were two questions about the share of water in certain vegetables. Questions related to diverse topics and each question was already tested in previous experiments (Lorenz et al., 2011; Rauhut and Lorenz, 2011; Moussaïd et al., 2013).[8]

In the decision phase II, each question had to be answered six times in a row, in a sequence of six rounds. After each round, participants received feedback about the answers and confidence statements provided by their group members according to a star network. The center of the star network could observe the previous answers and confidence levels of all four team members; the three pendants could only observe the previous answer and confidence of the center, in addition to their own. For each question of phase II, only one of the six rounds was selected at random by the end of the session to be payoff-relevant. Hence, there was no possibility to "hedge" risk with a portfolio of answers.

The actual treatments differed by the procedure that determined who within a group of four became the center of the star network for phase II. In the baseline treatment T0, the center was selected at random. In the accuracy treatment T1, the center became the group member whose guess on the similar question in phase I was closest to the correct answer. In the confidence treatment T2, the center became the group member whose level of confidence for the guess on the similar question in phase I was highest. Potential ties

[7] A more detailed description of the experimental procedures can be found in Online Appendix C.

[8] The full list of questions can be found in Online Appendix C.

in accuracy or confidence were broken at random. Half of all groups played the random treatment (T0) for four questions and the accuracy treatment (T1) for the other four questions; the other half played the random treatment (T0) for four questions and the confidence treatment (T2) for four questions. When the network for one question was formed, the selection procedure was made transparent to the group members. In the selection phase I, subjects did not know how decisions in the selection phase could have an influence on the decision phase. Instructions for the first phase simply announced that there would be a second phase with another set of instructions. This precluded strategic behavior in phase I, e.g., to become the center or to avoid becoming the center in phase II. While the answers to the questions were strongly incentivized, the confidence statements were not directly incentivized. Hence, the statements of confidence in phase II can also be considered as a mere communication technology. As we discuss in the next section, among rational agents there are indeed incentives to communicate truthfully the level of confidence in our setting in order to foster optimal learning in the group. However, our experimental results will not rely on the assumption that the confidence statements are truthful.

3 Theoretical Background

In this section, we derive theoretical predictions about the behavior in our experiment. The set-up is as follows. Let $N = \{1, 2, 3, 4\}$ be the agents in one team. Let 1 be the center of the star network and $2, 3, 4$ the pendants. The basic task in our experiment is to provide guesses on a specific question, the answer of which is a fraction. There is an unkown state of the world $\theta \in \Theta$, which is the correct answer to the question at hand.[9] Denote by $x_i(t)$ the answer of agent i at time t. Denote by $c_i(t)$ the confidence statement of agent i at time t. Time is discrete: $t = 1, 2, ..., T$, with $T = 6$ in phase II of the experiment. Accurate guesses are incentivized by a payoff function $\pi(e_i(t))$ that is weakly decreasing in the distance to the true answer $e_i(t) = |\theta - x_i(t)|$. One out of six answers is finally drawn as payoff-relevant.

To make predictions about the participants' guesses in phase II, we use two approaches: a rational learning approach and a naïve learning approach.

3.1 Rational Learning Approach: Bayesian Updating

In the rational learning approach, we assume that agents maximize expected payoffs given their beliefs and that beliefs are formed by Bayes rule.

[9] In the experiment, the correct answer is rounded and belongs to the finite set $\Theta = \{0, 0.01, 0, 02, ..., 0.99, 1\}$, which we can also model as the interval $\Theta = [0, 1]$.

Notice that a belief about the true answer is not a single number, but a probability distribution over the possible states ($f_i(t) : \Theta \to \mathbb{R}$). In the first round of guessing, $t = 1$, agents are endowed with some private information, i.e., what they know about the question at hand before interacting in the team. In the second round, each pendant $i \neq 1$ has observed the guess $x_1(1)$ and the confidence statement $c_1(1)$ of the center and can use this to update his belief. The center, on the other hand, has observed all guesses and confidence levels of the first round to form her belief, which is the basis for her second-round guess $x_1(2)$. If we assume that the guess and confidence level are sufficient to reconstruct an agent's belief and that the agents know how their private information is interrelated, then the center is fully informed after the first round of guesses. In this case, she can make the optimal guess $x^* := \arg\max_{x \in \Theta} E[\pi(|\theta - x|)|f_1(1), ..., f_4(1)]$, given the pieces of information in the team. Since all agents have the same payoff function and pendants can observe the center's guess $x_1(2) = x^*$, all agents make the same guess $x_i(t) = x^*$ from round 3 on. This observation leads to the following prediction.[10]

Prediction 1 (Bayes)**.** *In a model with common knowledge of rationality and common priors, the following holds. If the answer and confidence statement of a linked team member in a star network is sufficient fully to represent her private information, then the center learns once and the pendants learn twice. (Learning refers here to information updates and improvements in expectations.) Moreover, all team members will state the optimal answer x^* in any round $t \geq 3$, independent of who is at the center of the star network.*

Prediction 1 states that the selection of the team leader does not matter for the performance of social learning, apart from the first two rounds (and, in fact, only apart from round two). Moreover, it states that every agent states the payoff-maximizing guess, which implies that social learning is "efficient" in the sense of maximizing the sum of expected payoffs. However, several of its underlying assumptions deserve further attention.

First, a rational agent i is assumed to state the answer $x_i(t)$ that maximizes expected payoff, given his belief. This holds at least in the last round $t = 6$. In earlier rounds, there is potentially a strategic incentive to provide an answer that does not maximize expected payoff of that round (in order to be able to provide a better answer in later rounds). In fact, the earliest possibility to realize a deviating strategy is to deviate in round $t = 1$, learn something about the reaction of others in round $t = 2$, and materialize the better guess in round $t \geq 3$. Since, under the assumptions above, each agent states the optimal answer from round $t = 3$ on, strategic misrepresentation cannot pay off. There is simply no room for improvement. The same argument applies to the strategic misrepresentation of confidence statements. Hence, strategic misrepresentation is not an issue in our setting.

[10] A formal statement of this result can be found in Online Appendix B. There we introduce the general framework (B.1.1), prove the proposition (B.1.2), and provide two specific examples how such a rational model unfolds in our setting (B.2.1).

Second, it is explicitly assumed that statements of guesses and confidence levels are sufficient to recover beliefs. For this to be satisfied, the agent must know the other's belief up to one or two parameters. This is satisfied, for instance, in models assuming that beliefs follow a beta distribution.[11] Bayesian models with weaker assumptions could assume that agents also have beliefs about the signal quality of the others and imperfectly learn over time both the available private signals as well as their quality. Given the result by Aumann (1976), such a model is expected to lead to more learning iterations, but to the same outcome in the long run.

Third, how exactly an agent updates depends on his higher order beliefs on how private pieces of information are related to each other and how they are related to the truth. In theoretical models, it is usually assumed that there is common knowledge about the prior distribution of the true state, and about how private signals are drawn. In this experiment, agents are confronted with real questions. Hence, the agents' higher order beliefs about their own and their fellow team members' expertise can also depend on additional factors, such as the particular question at hand or on the treatment. In particular, the accuracy treatment T1, i.e., that the center gave the most accurate answer to a similar question, or the confidence treatment T2, i.e., that the center was the most confident on a similar question, might reveal something about the agent's ability that could be considered in the updating process. If anything, the declaration of the treatment T1 or T2 can reveal additional information, which would lead to better guesses, compared to the random treatment T0. To generate a prediction that is much more in line with the theoretical models, Prediction 1 abstracts from this possibility by assuming that there is common knowledge about how the private pieces of information are related to each other and to the truth.[12]

Fourth and finally, the assumption of common knowledge of rationality need not be satisfied. In sum, it cannot be expected that the requirements of Prediction 1 above are fully satisfied in the experiment. Still, the Prediction 1 gives us a clean baseline to compare the data with.

3.2 Naïve Learning Approach: DeGroot Model

Previous experimental research on social learning has not always found strong support for Bayesian learning, but often suggests that simple rules of updating, such as repeatedly taking averages, fit the data well (Corazzini et al., 2012; Grimm and Mengel, 2016; Battiston and Stanca, 2014; Chandrasekhar et al., 2016). We use their common modeling approach, which is often named after Morris DeGroot, to generate an alternative

[11] We study such models in Section B.2.1.

[12] In the experiment, we did not induce a common prior because we used questions of real topics. Nevertheless, we argue that models that assume a common prior and signals can contribute to our understanding of social learning in real settings.

prediction and to later specify models of more naïve learning. The basic aspect of naïveté incorporated in this modeling approach is that agents do not sufficiently account for the origin of information such that pieces of information are used each time they reach an agent through the network. This behavioral bias is also called "persuasion bias" (DeMarzo et al., 2003).

In the DeGroot model, the way people average the former guesses in their network neighborhood is typically constant. In the star network, this means that peripheral agents always provide a guess that is a mixture between the center's and their own last guess, with constant weights g_{i1} and g_{ii} on the two, while the center mixes all answers with some constant weights $g_{11}, g_{12}, g_{13}, g_{14}$, which are also positive and sum up to one. Given the weights and the initial answers $x_i(1)$, all consecutive answers $x_i(t)$ are fully determined. In particular, if G denotes the (row-stochastic) 4×4 matrix consisting of these entries g_{ij} and zeros at the remaining entries, the agents' updating can be written in vector and matrix notation as $x(t) = Gx(t-1)$. Hence, the predicted guesses are $x(t) = G^{t-1}x(1)$, for $t = 1, 2, ...$ Each agent thus generically changes guesses from round to round. Assuming that averaging weights are strictly positive is sufficient for the conclusion that all agent's guesses $x_i(t)$ converge for $t \to \infty$ to the same answer, which we denote by $x_i(\infty)$. Given that convergence is fast enough, $x_i(\infty)$ is also a good prediction for $x_i(6)$. It can be shown that, for any i,

$$x_i(\infty) = \frac{1}{c}\left(1x_1(1) + \frac{g_{12}}{g_{21}}x_2(1) + \frac{g_{13}}{g_{31}}x_3(1) + \frac{g_{14}}{g_{41}}x_4(1)\right), \tag{1}$$

with $c = 1 + \frac{g_{12}}{g_{21}} + \frac{g_{13}}{g_{31}} + \frac{g_{14}}{g_{41}}$. The weights $w_i = \frac{1}{c} \cdot \frac{g_{1i}}{g_{i1}}$ measure long-term influence of an agent i, which is called eigenvector centrality in network science since $w'G = w'$ (e.g., Friedkin 1991, DeMarzo et al. 2003, Golub and Jackson 2010). As can be directly observed from Equation 1, the center's influence on the long-term answer is different from a pendant i's influence, as long as $\frac{g_{1i}}{g_{i1}} \neq 1$. In particular, the center has a stronger influence if the center's weight on the pendant g_{1i} is lower than the pendant's weight on the center g_{i1}. This is a realistic assumption since pendants have only the center's guess to update from, while the center can distribute her weight among three pendants.

To discuss performance of social learning in this model type, we need to make assumptions about the relation between the initial guesses $x_i(1)$ and the truth θ, e.g., that initial guesses are realizations of independent random variables that have the truth as expected values. For any such probabilistic model and for any definition of the "optimal" guess $\hat{x}$ given the initial guesses, the approached value $x(\infty)$ and the optimal guess $\hat{x}$ will only coincide if by coincidence the averaging weights happen to be optimal in that sense. The same holds true for the guesses and optimal guesses of early rounds, say round two. Even if the weights g_{ij} happen to produce the optimal guess $\hat{x}$ for some agent i in some round t, they will not have this property for every agent and for every round. Hence,

there is an inherent inefficiency in these naïve models of social learning. The reason is that initial guesses of some participants are incorporated in the change of answers more frequently than other team members' guesses, while guessing weights are constant. These observations lead to the following prediction.[13]

Prediction 2 (DeGroot)**.** *In the naïve model with constant and positive averaging weights, the following holds. In a star network, every agent's learning heavily depends on the network structure, i.e., on who is the center. In particular, for $g_{i1} > g_{1i}$, the center has a larger influence on the long-run opinion than team member i. Generically, the center updates more than once and the pendants update more than twice. Under weak conditions, the first round of updating is learning (the expected error decreases), but for every notion of what is the optimal answer, the team members will generally state suboptimal answers.*

Prediction 2 states that the selection of the team leader heavily affects the performance of social learning, and that social learning is generally "inefficient" in the sense of not maximizing any function that is decreasing in the error of an agent's guess. Given the weighting matrix G, the naïve model is fully specified and provides a clear-cut prediction about all agents' guesses in all rounds. Typical specifications of G are studied in Section 5.3.

Our treatments T1 and T2 mainly affect naïve social learning through the manipulation of the network structure (who is at the center), but potentially also through the declaration of the treatments. The second channel would be present if the averaging weights g_{ij} depended on this declaration. In the empirical analysis, we will disentangle the effects of the manipulation of the center – which does not matter according to Prediction 1, but is crucial according to Prediction 2 – from potential effects of declaration (which can only be helpful in the rational framework of Prediction 1, but could also be harmful in the naïve framework of Prediction 2).

4 Success of Social Learning

The two theoretical approaches lead to contradicting predictions. Therefore, it remains an empirical question whether and how the selection of the leader affects the success of social learning.

4.1 Performance over Time

We measure the quality of the final answers both on the individual and on the collective level. On the individual level, we measure the quality by the error $e_i(t)$, which is the

[13] A formal statement of this result can be found in Online Appendix B.1.3. There we introduce a probabilistic framework and prove the proposition.

absolute distance between answer $x_i(t)$ and truth θ. On the group level, we use two complementary measures. First, we measure the quality of the four answers by the *collective error*, i.e., the error of the mean of the four answers in the group $ce(t) = |\frac{1}{4}\sum_{i=1}^{4} x_i(t) - \theta|$. Indeed, given the four final answers by a group, a decision might be taken on the basis of the mean of the four answers. Second, we consider whether the correct answer lies within the interval that is spanned by the four answers, and if so, whether it also lies within the interval that is spanned by the two answers which are contained in the interval of the two other answers. We define the indicator variable *(wisdom of) crowd error* as follows: $woce(t) = 0$ if at most two answers are strictly below or strictly above the correct answer; $woce(t) = 1$ if three answers are strictly below or strictly above the correct answer; and $woce(t) = 2$ if the correct answer lies strictly above or below all four answers in the group. The crowd error measures the error made when assuming that the correct answer lies between the given answers. For all three measures of performance, smaller errors mean higher performance.

Figure 1 depicts the levels of these performance measures over time, distinguishing by the three treatments. Panels A-C show that the individual errors are on average between 10 and 20 percentage points from the true answer and tend to decrease over time. More precisely, the pendants' average error reduces three times significantly on the five percent level. As intended in the accuracy treatment T1, selecting a center who was most accurate in answering a similar question (in phase I) leads to centers who are significantly better in estimating the current question in the first round (of phase II). The centers' average error reduces significantly once in the random treatment T0, as well as in the confidence treatment T2, but never so in the accuracy treatment T1 (at significance levels $p < 0.05$). By and large, these observations on the learning dynamics are consistent with the predictions of the rational model, namely that pendants learn twice and centers learn once. In particular, in the random treatment T0 the center reduces her error drastically from the first round to the second without significant further improvements, as the rational model would predict. Panels D-F show that the collective errors are on average between 12 and 16 percentage points from the true answer and also reduce over time. Similarly to the individual errors, the collective errors first decrease and then seem to settle after a few rounds (at a point that is significantly greater than zero). Taking these observations on individual and collective errors together, agents do learn from each other, but most of learning takes place in the first and in the second round of updating, i.e., until round $t = 3$.[14] A similar pattern, albeit with a change of sign, can be observed in panels G-I for the crowd error: The crowd error increases over time

[14]Learning cannot stem from having more time to think about a question since participants of the experiment who are not confronted with any information about the guesses and confidence of others did not at all improve over time. We tested this possibility with participants of the experimental sessions who were not exposed to any information. We randomly selected these subjects from all participants of sessions whose number of participants was not divisible by four, the size of our groups.

with most of its changes until round $t = 3$. This observation is consistent with findings of Lorenz et al. (2011), who show that the exchange of opinions reduces the wisdom of crowds. Crowd error is an indicator variable of which the averages have to be interpreted correspondingly. For instance, in the random treatment T0, *woce*(6) is 1.57 on average, which indicates that there are many cases (here: 65.9%) with a crowd error of two, and very few cases (8.5%) with a crowd error of zero. Hence, in the final period the correct answer most frequently lies outside of the convex hull of the provided answers.

Result 1. *Individual and collective errors reduce over time. Centers learn once (except in the accuracy treatment T1); pendants learn at least twice. Crowd errors increase over time.*

4.2 Treatment Effects on Performance

To test for treatment effects, we run regressions with the three error measures as the dependent variables and with treatment dummies as the independent variables. We focus our analysis on investigating the effects of learning on the final period, which is period 6. The last period is the most relevant, since it is the last period up to which learning can take place. In consecutive robustness analyses, we also analyze performance for earlier rounds back to period $t = 3$, the first round in which full learning can theoretically take place. Notice that the distribution of (individual and collective) errors is heavily skewed. Taking the logarithm (e.g., $\log(e_i(t) + 1)$) in the regressions of individual and collective errors gives less weight to errors which are far away from the truth and more weight to errors close to the true answer, such that the analysis will not be driven by a few cases in which errors were huge, say, forty and more. For the variable crowd error, which may attain values 0, 1, and 2, we use ordered logit.

Table 1 reports these models when controlling for each treatment T1 and T2 with a dummy variable, while T0 is the reference category. We control for the heterogeneity between different questions by using dummy variables. If selecting the most accurate or the most confident enhances performance, then we should see a significant negative effect on the three errors. As Table 1 reveals, the accuracy treatment T1 does not outperform the random treatment T0. The coefficients are insignificant and even positive. Even more strikingly, the confidence treatment T2 underperforms compared with the random treatment T0. The latter effect is significant on the 5-percent level for the individual error and the crowd error, and significant on the 10-percent level for the collective error.

Result 2. *Performance does not improve when the center is known to be the most accurate (T1). Performance even deteriorates when the center is known to be the most confident (T2).*

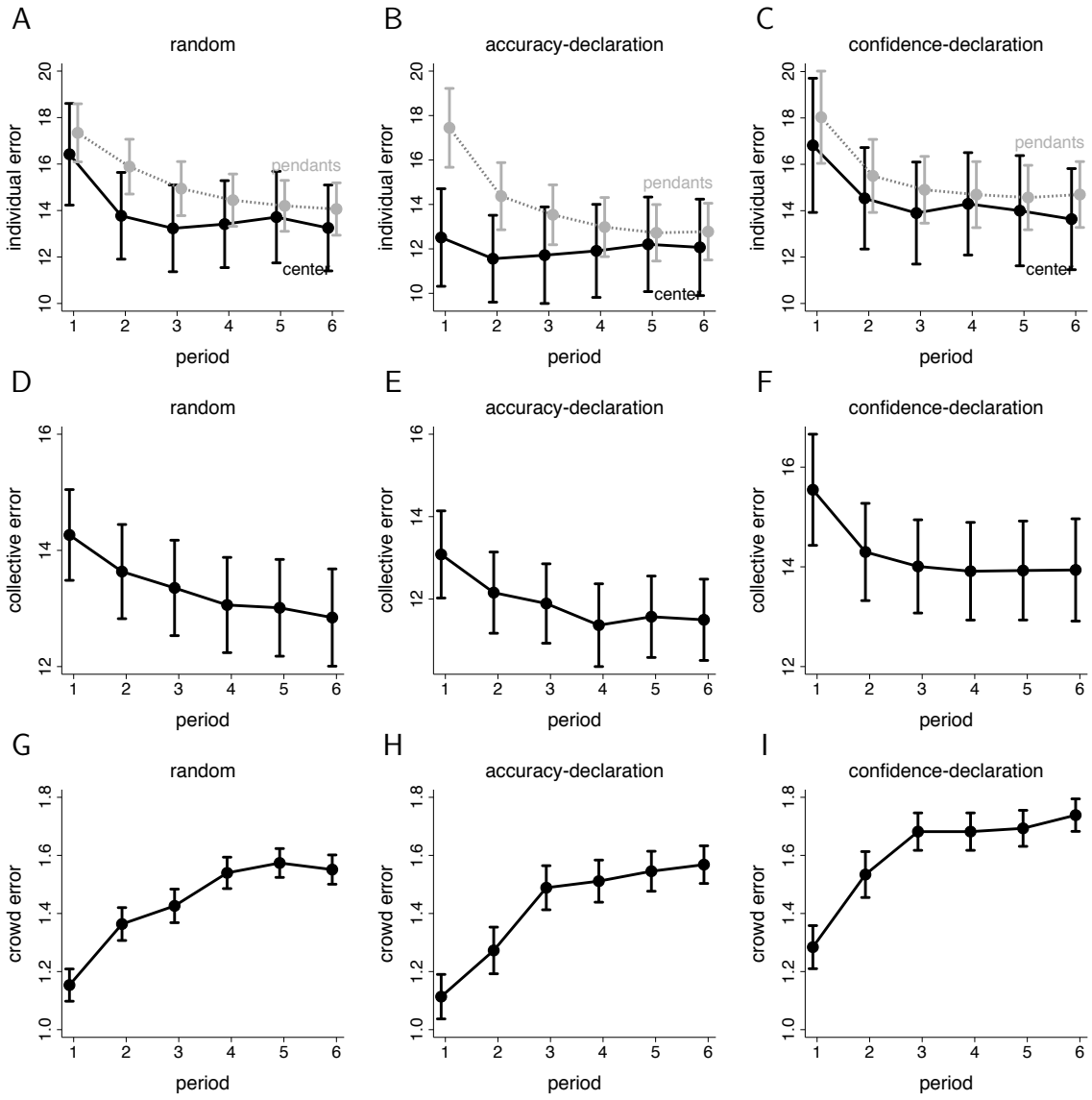


Figure 1: Individual, collective, and crowd errors over time by treatments. Panels A, B, C differentiate between centers (black) and pendants (gray). All confidence intervals are standard 95% confidence intervals.

	(1) individual error (log)	(2) collective error (log)	(3) crowd error
accuracy treatment (T1)	0.026	0.003	0.106
	(0.49)	(0.03)	(0.40)
confidence treatment (T2)	0.144*	0.179	0.739*
	(2.44)	(1.80)	(2.39)
intercept	2.164***	2.149***	
	(33.40)	(17.77)	
intercept cut 1			-2.555***
			(-6.76)
intercept cut 2			-0.830*
			(-2.51)
N	1.408	352	352

t statistics in parentheses

Question dummy coefficients for 8 questions not shown

Individual error: robust s.e. clustered for 176 subjects

Collective and crowd errors: robust s.e. clustered for 44 groups

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 1: Treatment effects on final errors: log error, log collective error, and wisdom of crowd error (in period 6). Linear regression (models 1 and 2) and ordered logit regression (model 3).

To understand the mechanism behind these treatment effects of selecting the most accurate or the most confident agent as a center, we distinguish between two aspects of each treatment, the trait of the central agent and the declaration of how the central agent was selected. By our experimental design we can disentangle the two effects, since in the random treatment T0 it frequently happens by chance that the most accurate agent was selected as the center without having the declaration of her or his accuracy, as is the case in the T1 treatment. The same applies for confidence; in a number of cases, the most confident agent was randomly selected to be the center in the random treatment T0.

Table 2 reports the results of the regressions when we control for the trait that the center is the most accurate or the most confident in the group, such that the treatment dummies only pick up the declaration effect. When the center happens to be the most confident or the most accurate, the outcome measures tend to improve, which can be seen from the negative sign of the (non-significant) coefficients. When the confidence of the center is declared to all group members, however, the performance is significantly reduced. The results are qualitatively similar for accuracy of the center in the sense that the signs of the effects are the same, but we cannot reject the null in that case, and the size of the effects is also smaller than for confidence.

While Table 2 reports the effects for the final period after all learning has taken place, Figure 2 illustrates robustness analyses of declaration effects when the regressions are run for each period separately. We show periods 3 to 6, since these are the periods after which full learning could happen and did take place according to the error dynamics (Figure 1).

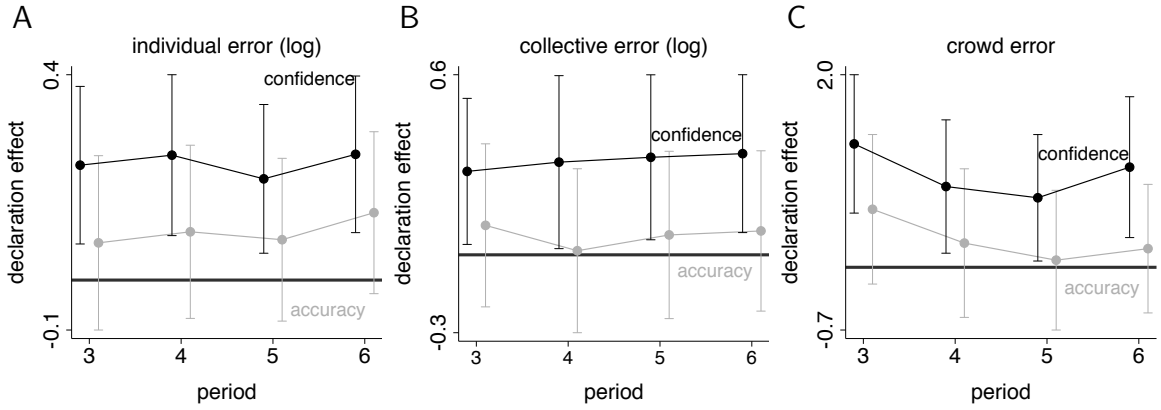


Figure 2: Treatment effects on errors: log error, log collective error, and wisdom of crowd error (periods 3-6). Linear regressions, 95 % confidence intervals.

The effect of declaring that the center is the most confident consistently increases the error measures and thus reduces performance. The declaration of accuracy has the same tendency, but the effects are smaller and insignificant.

	(1) individual error (log)	(2) collective error (log)	(3) crowd error
accuracy-trait	-0.110	-0.0716	-0.0477
	(-1.88)	(-0.76)	(-0.15)
accuracy-declaration (T1)	0.117	0.0790	0.196
	(1.64)	(0.60)	(0.57)
confidence-trait	-0.106	-0.231*	-0.474
	(-1.95)	(-2.20)	(-1.74)
confidence-declaration (T2)	0.218**	0.335*	1.053**
	(3.17)	(2.58)	(2.79)
intercept	2.221***	2.241***	
	(34.17)	(20.06)	
intercept cut 1			-2.735***
			(-6.99)
intercept cut 2			-0.999**
			(-2.89)
N	1.408	352	352

t statistics in parentheses

Question dummy coefficients for 8 questions not shown

Individual error: robust s.e. clustered for 176 subjects

Collective and crowd errors: robust s.e. clustered for 44 groups

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2: Treatment effects on final errors: log error, log collective error, and wisdom of crowd error (in period 6). Linear regression (models 1 and 2) and ordered logit (model 3).

Result 3. *Performance tends to improve when the center is the most confident. Declaration of confidence undermines performance.*

4.3 Social Influence

To analyze why the selection of the center can have a negative impact on performance, we study to which extent agents within a group influence each other. For this purpose we regress the answer $x_i(t)$ of an agent i in time $t \geq 3$ on his initial answer $x_i(1)$, as well as on the initial answers of the other group members $x_j(1)$. In particular, a pendant's answer is regressed on the center's initial answer, his own initial answer, and the mean of the other two pendants' initial answers. The center's answer is regressed on the average of the pendants' initial answers.

Tables A.1 and A.2 in the Appendix report the influence weights when estimating them separately for each treatment. For instance, in the random treatment T0, a pendant's final answer is estimated as the linear combination of its initial answer with weight 56.7%, the center's initial answer with weight 26.7%, and the other pendants' average initial answer with weight 16.6%. There are several interesting observations to make in these tables. First, every agent places much weight to his own initial opinion. In the rational model and the random treatment, we would expect that on average this weight is 25%.[15] Second, the weight individuals place on their own initial opinion depends on the treatment. In the random treatment, pendants place more weight on themselves than in the other two, while centers place less weight on themselves in the random treatment. Finally, the social influence by the other team members heavily depends on the treatment. For pendants, the center's weight was 26.7% in the random treatment T0, but 46.9% in the confidence treatment T2; and similarly in the accuracy treatment T1.

The two aspects of a treatment, the trait of the center and the declaration of the center, are then captured by the interaction effects of the corresponding dummy variables with the influence weights in the regressions that pool the three treatments. These regressions are reported in Tables A.3 and A.4 in the Appendix. Their effects are illustrated in Figure 3. A positive effect of a certain dummy variable means that the given influence weight is increased by the given treatment.

When the center happens to be the most accurate or the most confident, but there is no public declaration of this, then the pendants do not strongly respond (panel A); they only mildly increase their weight on the center. In the same case, i.e., when the center is the most accurate or confident, the center places significantly more weight on her own initial opinion and, accordingly, significantly less weight on the pendants' opinions (panel B). In contrast, the declaration that the center is the most confident or accurate does not affect the center's weighting (panel D), but there is a strong effect on the pendants

[15] We will return to this observation when extending the social learning models in section 5.1.

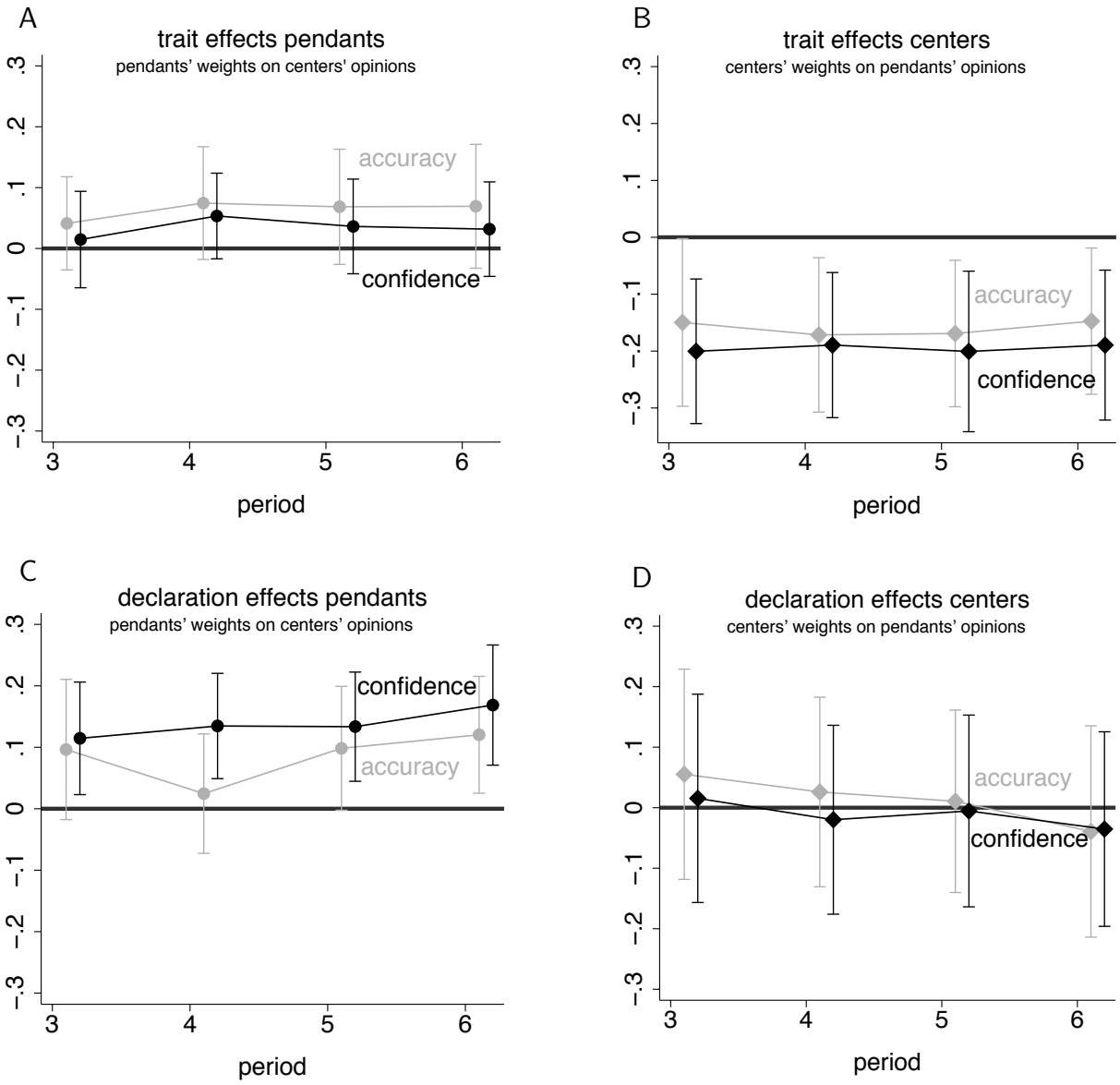


Figure 3: Trait and declaration influence for pendants and centers. Gray accuracy, black confidence treatments, 95 % confidence intervals.

(panel C). Declaring that the center is somehow special (the most confident or accurate on a similar question) significantly increases the pendants' weights on the center's initial opinions.

Result 4. *The pendants place more weight on a center who is* declared *to be the most confident or the most accurate. The center places less weight on the pendants' when she is the most confident or the most accurate.*

This result provides an explanation for the former results. Declaring that the center is somewhat special increases the weight (s)he receives. Placing more weight to a single opinion has a negative effect on performance, except if this person is substantially better informed than the others. In the accuracy treatment T1, this condition is satisfied to some extent, such that the negative effect of placing too much weight on a single person and the positive effect of placing more weight on a person who is better informed may balance each other. Consequently, the performance in the accuracy treatment T1 need not differ from the random treatment T0. In the case of the confidence treatment T2, the center is not substantially better informed than the other group members, as can be seen from panel C in Figure 1. Hence, giving him/her more weight only has the negative effect of insufficiently taking into account the information of the others. This is on average even worse than the random treatment T0.

4.4 Overconfidence

As we have seen in Table 2 above, it is rather beneficial for the group when the center happens to be most accurate or most confident, but is not declared as such. On the other hand, it is well-known that many people are often overconfident, i.e., they report much too small confidence intervals when asked about a region where they expect the true answer with a certain probability (a usual way is to ask where they expect the answer in 90% of their guesses; see, e.g., Soll and Klayman (2004); Moore and Healy (2008); Herz et al. (2014). In phase I of our experiment, we asked participants to provide such regions. Therefore, we can compute for every participant her individual overconfidence score simply by counting how often that person provided a confidence interval that did not contain the true answer. Thus, every participant is characterized by an overconfidence score in $\{0, 1, \ldots, 8\}$ with the interpretation that a person is the more overconfident the larger her overconfidence score becomes. As Figure 4 reveals, many agents are overconfident. Their guess should only lie in 10% of the cases outside of their provided 90% confidence interval. However, for most agents this happens in more than two out of eight cases. The histogram also documents that there is substantial heterogeneity in overconfidence.

In Table A.5 in the Appendix, we analyze how the center's overconfidence score as well as the sum of the pendants' overconfidence scores impact the group's performance

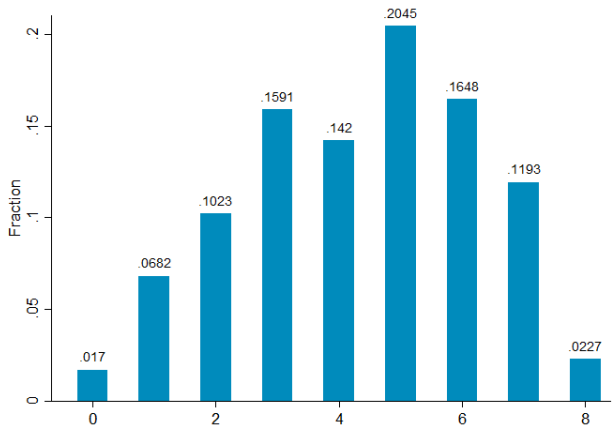


Figure 4: Histogram of overconfidence. The value 0 means that a subject has specified for all eight knowledge questions a respective 90% confidence interval which encloses the true value. The value 8 means that a subject has specified for all eight knowledge questions a 90% confidence interval which does not enclose the true value. All values above 1 indicate overprecision, since more than 10% of estimates fall out of the 90% confidence interval (i.e., 91.5% of subjects are overconfident).

(on top of the previously found treatment effects): we find the corresponding regression coefficients to be significantly positive for all three error measures (while the formerly discussed effects remain). Moreover, the center's coefficient is substantially larger than the pendants'. We thus have the following result.

Result 5. *Both the center's and the pendants' overconfidence undermine performance. The center's overconfidence has a more deteriorating effect than the pendants' overconfidence.*

Given this result, it is, ceteris paribus, best for the group's performance if the most overconfident group members are pendants, i.e., it would be best if the least overconfident group member was the center. On the other hand, overconfidence is of course related to confidence itself, and the most confident group member acting as center improves the group's performance when she is not declared to be the most confident. Indeed, Table A.5 reveals that, when controlling for overconfidence and for the declaration of confidence, the trait of being the most confident significantly increases performance.

Thus, we conclude that the leader personality who should optimally be selected is characterized as confident without being overconfident. Depending on the individual characteristics, it can therefore be optimal to select someone who is not the most confident agent, if the chosen agent can compensate by being very 'tight', i.e., not overconfident.

Hence, all results (Results 1-5) contribute to a coherent picture of how the selection of the leader affects social learning. To investigate this interpretation further, in particular

the one of placing "too much weight on the center" in the confidence treatment T2, we analyze more in-depth the underlying micro-level mechanisms. In particular, we will study the fact that weights on own opinions are too large, which also prevents optimal social learning. As the social influence analysis showed, both pendants and centers generally placed much weight on their own initial opinion. When studying the learning behavior in the next section, we will incorporate this behavioral aspect.

5 Learning Behavior

The experimental data allow us to test theories of social learning on multiple levels. First, their implications for the performance of social learning (as summarized in Prediction 1 and Prediction 2) are found to be consistent with some empirical results and inconsistent with others. Second, we can directly take the theoretical models to the data and study which aspects are in line with real behavior. For this purpose, we specify and vary the models and measure which model specification best fits the data. We thus include model variations that incorporate conservatism, a pattern that is commonly found in experimental set-ups, but absent in any Bayesian model of social learning (that we are aware of) and absent in almost all naïve models of social learning.

5.1 Specification and Extension of Bayesian Models

To specify the rational models, we assume that each agent's belief follows a beta distribution. This is a standard functional form for beliefs that live on intervals.[16] With some assumptions on the distribution of signals, all agents' beliefs at any time indeed belong to the class of beta distributions.[17] Assuming conditional independence of initial signals, Bayesian agents will state guesses that are convex combinations of their initial guesses. The weight on these guesses, however, depends on the signal quality of each agent i, which we denote by n_i. The model variations that we study differ in the assumptions about signal quality.

A baseline assumption is to suppose that the precision of each agent's signal is the same, i.e., $n_i = n_j$ for all i, j. In that case, the optimal guess x^*, which will be the consensus from round $t = 3$ on, is simply the unweighted mean of the initial guesses $x_i(1)$. We call this the *Standard Model*. Alternatively, agents are assumed to communicate their belief fully by providing the guess and the confidence level. Then, for each answer $x_i(1)$ and its confidence $c_i(1)$, the center can determine the two parameters of the corresponding beta distribution and combine all initial beliefs in a rational manner, thereby updating

[16] Like the normal distribution, which is a standard functional form for beliefs on the unbounded real numbers, it is determined by two parameters only.

[17] The formal framework is provided in section B.2 of the Online Appendix.

leads to a combination of own and others' guesses – not with equal weights, but with larger weights for those guesses which are tagged by high confidence. We call this the *Sophisticated Model*. Note that these are two opposing views on the informativeness of the confidence statement – either confidence is fully informative or confidence can be ignored – which lead to two models that both satisfy the requirements of Prediction 1, and are hence similar in most respects. They differ in their weighting of initial information.

The previous empirical literature on real people's beliefs and their updating finds two very strong and consistent patterns: overprecision and conservatism.[18] There is a simple way to introduce both of them into our model: Agents overestimate their own signal precision by a factor $\tau_i \geq 1$; respectively, they underestimate the signal precision of the others by the inverse factor $\frac{1}{\tau_i}$. The motivation of this model variant is that overconfident agents suffer from overprecision in the sense that they perceive their signal as more precise than it is.[19]

Formally, this is a generalization of the *Standard Model* and the *Sophisticated Model*. This model also predicts that there are no more changes after $t = 3$. However, this model does not predict consensus! The agents' opinions settle down in between the prediction of x^* (i.e., the case $\tau_i = 1$ for all $i \in N$) and their initial guess $x_i(1)$. The weight of the own initial guess is thereby increasing in overprecision τ_i. In particular, if $\tau_i \to \infty$, then $x_i(t) \to x_i(1)$, i.e., infinitely overprecise agents are totally conservative and always stick to their initial guess. (We will include such a model as a baseline and call it the *Sticking Model*.)

To specify concrete models, we choose levels of overprecision τ_i that match with empirical results on overprecision. When asked for a 90% confidence interval, many people provide a 50% confidence interval instead. This is roughly induced by $\tau_i = 5$. Incorporating conservatism of every agent into the *Standard Model* or, respectively, into the *Sophisticated Model* leads to the two models *Standard-Plus Model* and *Sophisticated-Plus Model*. In the *Standard-Plus Model*, agents behave very similarly to the *Standard Model*, but move only a fraction into the direction of the center, which corresponds to findings on conservatism. The only difference to the *Sophisticated-Plus Model* is simply that we specify the initial signal precision not as equal, but according to the confidence statements. Agents are assumed to know that others are overprecise and thus learn about the original signals by correcting for τ.[20]

[18] Overprecision, as it is called by Moore and Healy (2008), is also known as "judgmental overconfidence" (Herz et al., 2014), "overconfidence in interval estimates" (Soll and Klayman, 2004), or "resoluteness" (Bolton et al., 2013), and is defined as "excessive certainty regarding the accuracy of one's belief." Conservatism means that agents are not willing to learn sufficiently from new signals (e.g., Peterson and Beach (1967); Mobius et al. (2011); Ambuehl and Li (2014); Mannes and Moore (2013)). Of course, the two patterns are closely related to each other.

[19] Or, alternatively: agents learn from their neighbors, but they attach higher uncertainty to the beliefs of others than to their own belief.

[20] In the conservatism models (consisting of the specifications Standard-Plus and Sophisticated-Plus),

Importantly, the four models *Standard Model*, *Sophisticated Model*, *Standard-Plus Model*, and *Sophisticated-Plus Model* are all special cases of Bayesian models and hence produce the prediction that is formalized as Prediction 1. Except that, in the *Standard-Plus Model* and the *Sophisticated-Plus Model*, agents do not state the same guess x^* from round 3 on, but their subjectively perceived optimal guess x_i^*, which is a mixture between x^* and the agent's initial guess $x_i(1)$. This difference is illustrated in Figure 5 below in the two left panels, which compare the dynamics of the *Standard Model* with the *Standard-Plus Model* in a simple example.

5.2 Specification and Extension of DeGroot models

In the DeGroot framework of naïve learning, agents approach consensus. Consensus is given by $x(\infty) = w'x(1)$, where the vector w captures the eigenvector centrality of the agents (e.g., Friedkin (1991); DeMarzo et al. (2003); Golub and Jackson (2010)).

The most common specification is to allocate equal weights to any connection including to oneself.

$$G = \begin{pmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} \end{pmatrix}$$

Credit for this specification is usually given to DeMarzo et al. (2003). This behavior corresponds to Bayesian updating with independent signals of equal precision in the first round, but not in later rounds. The long-term prediction using this *DeMarzo et al. Model* is determined by $w = (\frac{2}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5})'$, i.e., pendants' initial opinions enter the calculation of the consensus with a weight of 20% each, while the center's initial opinion accounts for 40% of the consensus.

Corazzini et al. (2012) suggest improving the *DeMarzo et al. Model* by increasing the weight of agents who listen to many other agents (and show that this twist improves the model fit to experimental data). The suggested specification is that the weights are

we make assumptions about higher-order beliefs that close the model in the sense that no agent will expect another agent to behave in a different manner than in the one observed. In particular, we assume that all agents think of all other agents as overprecise; and that all agents think that all agents think that all agents are overprecise. In that way, an agent i is not surprised that j discounts i's behavior from i's point of view (from a neutral point of view, j takes i's behavior as he should) and that j overvalues j's guess (from i's and a neutral standpoint).

proportional to the outdegree (i.e., the number of agents listened to):

$$G = \begin{pmatrix} \frac{1}{2} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \\ \frac{3}{4} & \frac{1}{4} & 0 & 0 \\ \frac{3}{4} & 0 & \frac{1}{4} & 0 \\ \frac{3}{4} & 0 & 0 & \frac{1}{4} \end{pmatrix}$$

This model predicts that the center of the star is even more influential in the long run: $w = (\frac{9}{15}, \frac{2}{15}, \frac{2}{15}, \frac{2}{15})'$.[21]

Incorporating conservatism requires a model extension. Friedkin and Johnsen (1990) provide a more general model of naïve learning. Initial opinions are determined by some exogenous conditions, which can always have an impact on an agent's opinion. Such a model has also been analyzed in Golub and Jackson (2012). To incorporate this aspect, we can simply let agents stick to their initial guess $x_i(1)$ to some extent α:

$$x_i(t) = (1 - \alpha_i) \cdot Gx(t-1) + \alpha_i \cdot x_i(1).$$

For $\alpha_i = 0$, we have the DeGroot model. For $\alpha_i = 1$, we have the simplest conceivable model: an agent makes an initial guess $x_i(1)$ and then sticks to it. This is a baseline model that we call the *Sticking Model*, as already mentioned when discussing totally overprecise rational learners.

If $\alpha_i \in (0, 1)$ for every agent i, then the model prediction is that agents move towards the others' guesses, but still rely on their initial guess. This is conservatism.[22] Interestingly, with this model variation, the updating process converges without reaching a consensus (for generic starting values).

We extend the *DeMarzo et al. Model* and the *Corazzini et al. Model* by the Friedkin and Johnsen (1990) framework and set the conservatism/overprecision parameter $\alpha = 0.5$. This leads to the *DeMarzo et al. Plus Model* and the *Corazzini et al. Plus Model*. In these models, agents do not approach consensus anymore. For instance, in the *DeMarzo et al. Plus Model*, the long-term guess of a pendant i is a convex combination of initial guesses with the following weights: weight $\frac{2}{9}$ on the center's initial guess, weight $\frac{1}{27}$ on other pendants' initial guesses each, and weight $\frac{19}{27}$ ($\approx 70\%$) on the own initial guess, which leads to different guesses of each pendant. This difference is illustrated in the right panels of Figure 5. The long-term weights of the *Corazzini et al. Plus Model* are comparable, but

[21] Grimm and Mengel (2016) propose another specification of the DeGroot weights. However, their extension does not lead to an additional prediction here because weights depend on the clustering coefficient, which is zero for all agents in the star network.

[22] Interpretations for the cause of conservatism include forms of overprecision or kinds of anchoring bias in which the initial guess serves as anchor and the adjustments to the others' guesses is limited by parameter α.

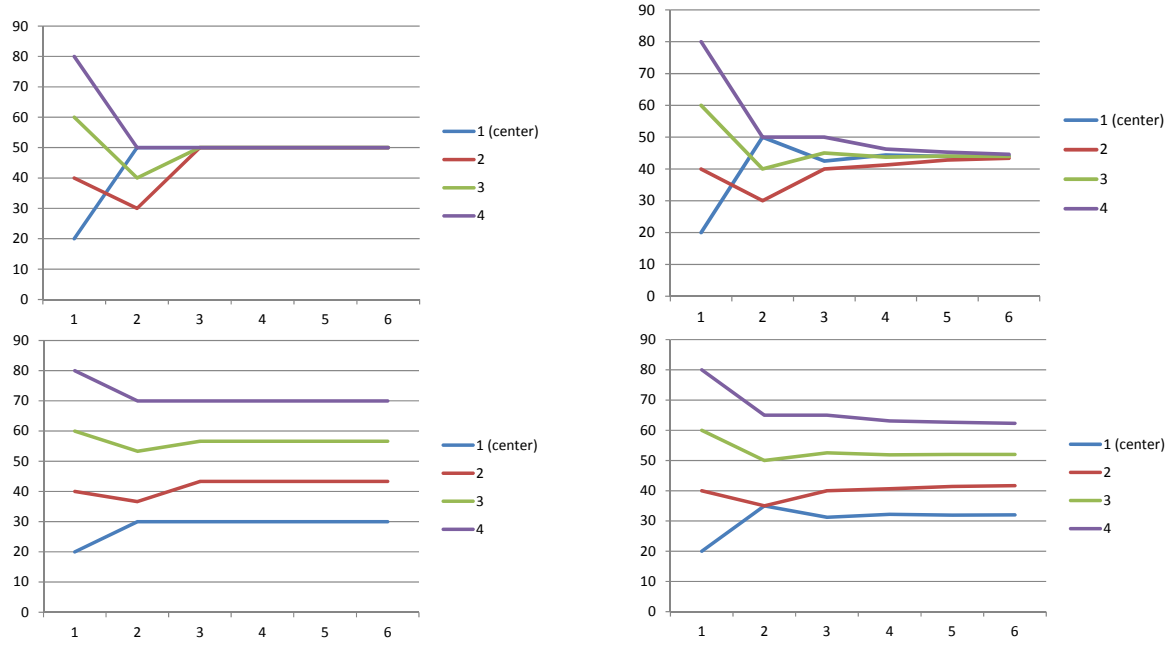


Figure 5: Simple examples of dynamics with time on the x-axis and answers (in percentage points) on the y-axis. Upper panels illustrate two prominent models from the literature; lower panels illustrate their extensions when conservatism is incorporated. *Standard Model* is upper left, *Standard-Plus Model* is lower left, *DeMarzo et al. Model* is upper right, and *DeMarzo et al. Plus Model* is lower right panel. Hence the left panels illustrate rational models, the right panels naïve models.

differ in that each agent, including the center, is more heavily influenced by the center's initial opinion.

Four models are illustrated in Figure 5. In this example, initial answers are $x_1 = 20\%$ for the center, and $x_2 = 40\%$, $x_3 = 60\%$, and $x_4 = 80\%$ for the pendants. The most important differences are easily observable. In Bayesian models (left panels), learning stops in round 3; in naïve models (right panels), answers converge. In the specifications without conservatism/overprecision (upper panels), agents reach or converge to consensus; in the models with conservatism/overprecision (lower panels), there is a persistent heterogeneity of answers, such that each agent's answer is "biased" towards the own initial answer. Note that the conservative/overprecise agents in the naïve models behave similarly to conservative/overprecise agents in the rational learning approach.

5.3 Comparison of Models (Horse Race)

In total, we have specified nine models. Four following from the rational approach to social learning, four following from the naïve approach to social learning, and one baseline

model (the *Sticking Model*), which is a degenerate special case of both model classes. We implemented each model such that all periods $t \geq 2$ are predicted from values at $t = 1$. We assess the fit of each model by measuring the root of the mean squared error (RMSE) between the model predictions for $t \geq 2$ and the data points. Figure 6 displays the results.

The worst overall model fit is obtained by the baseline model, in which all agents stick to their initial guess (*Sticking Model*). The best model fit is obtained by the "Plus" models, which incorporate conservatism. In fact, every model considered has a larger RMSE than its "Plus" counterpart that incorporates conservatism.

Considering the model fit for each round separately, the conservatism aspect seems particularly helpful in predicting the first updates (round 2). Hence, the "Plus" models fit much better than the others in these early periods. However, in the last period, the "Plus" models fit best still, with the only exception that the *DeMarzo et al. Model* fits better than the *Sophisticated-Plus Model*. This observation indicates that the advantage of the models that include conservatism is not restricted to the first rounds, but resists.

Ignoring the "Plus" models for one moment, we can see that the naïve learning in the DeMarzo specification fits well to the data. The sophisticated specification of the rational model does not fit to the data. The standard specification of rational learning and the Corazzini specification of the naïve learning are somewhere in between. Hence, the straightforward specifications that treat all agents symmetrically (*Standard Model*, *DeMarzo et al. Model*) are at least as adequate as the specifications that incorporate confidence statements in a specific way (*Sophisticated Model*), or that incorporate the unequal degree (*Corazzini et al. Model*).

Adding conservatism to the models leads to a very good fit of the rational model in its standard specification (*Standard-Plus Model*) and a better fit of the sophisticated specification (*Sophisticated-Plus Model*) than without conservatism. The best model fit is obtained for the naïve models with conservatism (*Corazzini et al. Plus Model* and *DeMarzo et al. Plus Model*).

We can also differentiate the model fit by treatment. The results are illustrated in Figure A.1 in the Appendix. The best model fit in the random treatment T0 is obtained for both the *DeMarzo et al. Plus Model* and the *Standard-Plus Model* with an RMSE of 7.88. Hence, these extensions of straightforward specifications of the naïve and the rational approach best predict the experimental data in the baseline treatment. Comparisons are similar across treatments. However, the *Corazzini et al. Model* fits better in the accuracy T1 and confidence treatment T2 than in the random treatment T0. The reason is that the center receives a high influence weight in the accuracy and confidence treatment T2, as well as in the *Corazzini et al. Model* specification. Complementarily, the baseline model of sticking to the initial guess fits much better in the random treatment T0 than in the others. This is a clear indication that social influence is weakest in the random treatment T0 and stronger in the accuracy treatment T1 and the confidence treatment T2. Given

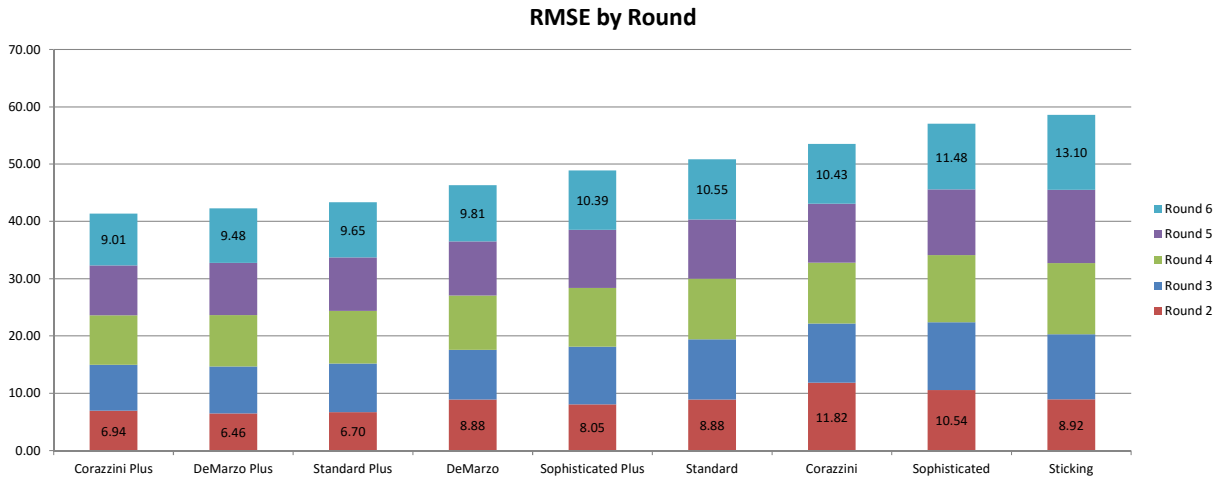


Figure 6: Root mean squared errors (RMSE) of social learning models. "Standard" and "Sophisticated" are models of rational learning; "DeMarzo" and "Corazzini" are models of naïve learning. "Plus" models incorporate conservatism. Lower errors mean better fit between model and data.

that social influence can undermine the wisdom of crowds (Lorenz et al., 2011), this is an explanation for our result that the crowd error is lowest under the random leader T0.

We finally differentiate between the model fit for the center and for the pendants. The result is displayed in Figure A.2 in the Appendix. The *Corazzini et al. Model*, which predicts an immense influence of the center, fits well for the center, but not for the pendants. Again, the "Plus" models fit well generally for both pendants and centers. The best fit for the pendants is attained by the *Corazzini et al. Plus Model*, and the best fit for the center is attained by the *Standard-Plus Model*.

Result 6. *Incorporating "conservatism" into both the rational and naïve models of social learning increases the fit between theoretical models and empirical data.*

The result holds for all four considered models, for all three treatments, for all rounds, and, apart from one exception, for both centers and pendants. The exception is that the *Corazzini et al. Model* predicts the center's opinion better than the *Corazzini et al. Plus Model*. Hence, our data strongly indicate that the extension of both the rational and the naïve models of social learning by conservatism is not a mere theoretical exercise, but an empirically relevant generalization.

In sum, the results of the horse race show, first of all, that both models of rational and models of naïve learning can contribute to our understanding of social learning in teams. Second, the baseline model that each agent sticks to his own initial guess and keeps his independent opinion fits much better to the data when the team leader was selected at random. Complementarily, models that predict an immense weight of the team leader's

opinion (*Corazzini et al. Model*) fit well when the leader is known to be the most confident (T2) or most accurate (T1).

Finally, the known models of social learning might fall short of covering the substantial amount of conservatism that is characteristic for the social learning of real people. Assuming that people are overprecise provides a foundation for conservative learning, even for rational learners, and affects the model prediction such that they are much closer to our data. We can connect this observation with Result 5 that overconfident leaders undermine social learning. Assuming that agents are overprecise induces conservative learning in which the opinions of others are not sufficiently accounted for. Therefore, overconfident leaders undermine performance.

6 Discussion

6.1 Summary and Conclusions

An organization's fit to the environment depends on the management's ability to assess the state of the – usually dynamic – environment and to cope with uncertainty. We measure team performance in this respect by assessing its ability to estimate correct answers to factual questions.

Having a team leader who is knowledgeable or confident in a given topic might in principle be helpful. However, communicating the leader's qualities can undermine this effect. Stressing the expertise or confidence of the leader triggers other team members to put too much weight on the leaders' opinion. This narrows the opinion space and diminishes the wisdom of the group substantially. Past accuracy (T1) and actual ability are correlated such that there is a positive effect of an accurate leader, which, however, is immediately undermined by the effect of declaring it. Confidence (T2) is only weakly correlated with actual ability such that the net effect is significantly negative.

In addition to a negative effect of declaring the selection criteria of leaders, we can show that most people are overconfident in their estimates and in their assessments of problems. Overconfidence leads to ignorance of the others' valuable opinions, information gets lost, and the team's potential for solving problems deteriorates. While overconfidence of every team member has a deteriorating effect, the leader's overconfidence has the strongest negative effect.

These are two detrimental effects of leaders selected by confidence. We can further show the micro-mechanisms of these detrimental effects by simulating different classes of learning models. In particular, rational learning models in which social learning is efficient, independent of the team leader, fall short of explaining our data. A better fit is obtained for naïve learning models that predict that the leader is more influential than any other team member. Among those, the model that gives tremendous weight to the leader

(*Corazzini et al. Model*) does not fit well in the random treatment T0, but particularly well in the treatments T1 and T2, in which the leader is not selected at random. Compared to all models, people tend to adapt too little to the others' opinions and are too confident in their own subjective estimates. To introduce this pattern in the theory of social learning, we extend both rational and naïve models by conservatism, which can be derived from overconfidence. With this twist, the fit of each model to the data increases substantially. Moreover, this kind of bounded rationality leads to the fact that leaders learn too little from the opinions in their network.

One conclusion from our paper is that we provide evidence for the superiority of a selection procedure that is based on random leader selection ("sortition"). This mechanism has its roots in ancient Greece and has been discussed by various names such as "demarchy" or "aleatory democracy" (Zeitoun et al., 2014; Frey and Osterloh, 2016). While there have been discussions in the literature about the advantages and disadvantages of aleatory democracy, there is hardly empirical evidence. Our empirical results demonstrate that random selection can be beneficial compared to selection based on confidence. Selection by confidence often leads to detrimental effects of truth-finding, since first the leader listens too little to other members of the group and second the other members listen too much to the leader. Our experiment is one of the first to shed light on one of the potential mechanisms of why aleatory democracy may be beneficial. The strength of random selection is not restricted to reducing the probability of overconfident leaders. It is also based on the fact that the leader's influence on team members is not amplified and, therefore, the others' opinions are respected more compared to a system in which leaders push their own views on all team members due to both a central position in the communication network and an additional legitimacy because they are selected by expertise, or even worse, by confidence.

The problem of overconfidence of leaders and its detrimental effects on group wisdom becomes even more important when considering that expertise is often hard to measure in reality. In fact, publicly expressed subjective confidence in the own expertise might sometimes be more important for becoming a leader than objective expertise. The problem is that the truth is often not precisely known. Therefore, publicly expressed confidence may persuade others that the person in question may know the correct answers. However, as our experiment shows, when confidence and expertise are not strongly correlated, overly confident leaders can mislead the group. This difficulty in assessing the true expertise for selecting leaders may therefore be another argument for the beneficial empirical effects of aleatory democracy, where the leaders are selected at random.

In our experiment, we focus on the team's ability to converge to correct assessments of the environment, which is to adapt and learn from each other such that they find correct answers to factual questions. However, in addition to correct problem-solving, another goal of teams is to foster cohesion, e.g., to strengthen their corporate identity. Sometimes,

it is less important to find the truth, but more important to converge towards a common opinion. Having a common opinion helps to reduce conflicts, work on the same tasks, and help each other. This means that opinion convergence can be a separate, distinct goal of teams and those leaders may be preferable who manage to unify the opinion space in their team. Our experiment only focuses on the goal of finding correct answers. How to foster coordination, opinion convergence, and cohesion is another goal. For example, it has been shown that a leader's overconfidence or resoluteness can foster coordination and cohesion (Bolton et al., 2013). In their theoretical contribution, Bolton et al. (2013) already point to the trade-off that an overconfident leader, while having positive effects for coordination, has the downside of not sufficiently learning from the followers. We can now strengthen and empirically document the second mechanism: overconfidence of leaders is clearly a detrimental factor to the team's learning. Hence, the strength of overconfident leaders for coordination comes at the downside of suboptimal information-processing. When it comes to tasks which are related to truth-finding, we claim that overconfident (or resolute) leaders are actually harmful.

6.2 Limitations

The strength of our experimental design comes at the expense of certain limitations. First, the external validity of this type of experiments depends on whether the interaction among participants (who were virtually all university students) are similar to the interaction among members of real teams in organizations. Moreover, we focus on the organizational task to estimate the state of the environment, while other aspects also matter for the performance of an organization. At the same time, this is a strength of our experimental set-up, since we can isolate the performance in a key task: estimating the state of the environment.

We have exogenously varied the selection criterion of the leader. This takes the perspective of the top management, deciding about, e.g., the promotion criteria of more and less senior employees of the organization. It would also be interesting to see how team members themselves would choose a selection criterion if they were given the opportunity to choose.

By studying star networks, we have not varied the network architecture, but only the network positions, which for star networks boils down to the question of who is the leader. Follow-up research might include a variety of network architectures. This is beyond the scope of this paper because it would shift the emphasis from the selection of the team leader to the selection of a communication architecture within an organization. Formal hierarchies within organizations usually have a star-like structure, e.g., they determine the head of an organizational unit, or the president of a certain committee, which can be directly modeled by star networks. However, since informal networks within organizations

are also known to be important, alternative network architectures and even endogenous network formation should be considered in future research.[23]

Finding that overconfidence is an important determinant of social learning suggests an alternative treatment that combines accuracy and confidence to overconfidence, in which leaders are selected based on their relatively low or high level of overconfidence. This seems an interesting extension that, however, does not match real selection procedures we are aware of. This could be considered an innovative suggestion to assess overconfidence when selecting managers. Our treatments T1 accuracy and T2 confidence resemble real selection criteria based on maximal competence, which are either objectively assessed (T1 accuracy) or subjectively provided by self-declaration (T2 confidence).

Finally, our experimental design focuses on social learning and does not mix it with the decision-making process. After learning took place in a team, there are various forms of how a decision is actually made. It could be the case that the team communicates its opinions to the higher level management or their client, who then draw their conclusions and take actions. It could also be that the team takes actions on its own, deciding, e.g., by the majority rule or with unanimity about the consequences. Obviously, decision-making processes also affect the quality of the decisions and are thus important to study. However, studying them jointly with the social learning process can distort the measures of learning since communication before collective decisions makes strategic considerations in the communication stage prevalent.

6.3 Practical implications

Our findings suggest several practical implications. First, when selecting a leader, there is a substantial difference between assessing a candidate's competence by some tests (as in our accuracy treatment T1) versus relying on her subjective statement of her own competence (as in our confidence treatment T2). This even holds when there are no strategic incentives to misrepresent the own opinion and the own competence. Our findings clearly suggest, whenever possible, focusing on objective measures of competence rather than trusting subjectively stated confidence in candidates' own expertise. A large majority of people is overconfident, such that starting a competition as to who is claiming the highest confidence will most likely lead to detrimental effects in selecting leaders who will listen too little to other opinions in their network. Hence, when the main goal of the team is related to truth-finding, this is expected to be a poor selection criterion.

Second, the way the selection criterion for the leader is communicated to a team heavily affects the team's interaction and performance. In particular, making explicit that the team leader was selected at random can lead other team members to make use of their

[23] In a quite different framework, endogenous network structures and social learing in organizations have been studied by Çelen and Hyndman (2012).

own valuable knowledge instead of "blindly" following their leader. By the hierarchical structure, which determines the communication network, the team leader is already very powerful and her opinion is certainly heard. Declaring that the team leader was selected because of her (alleged) superiority increases her power, which might push team learning out of balance. Hence, keeping quiet about the (alleged) superiority of a team leader can foster more efficient learning within a team.

Third, we can validate that communication and social influence can be harmful for the wisdom of crowds effect (Lorenz et al., 2011). We confirm this finding for unequal communication structures in terms of star networks, where all people are connected to a single center, who receives all information from the network while the others have to communicate via the center. In particular, we show that the wisdom of crowd error increases over time, giving evidence that the group can exploit less and less information from other network members over consecutive periods of social influence. However, and importantly, we also show that social influence can foster social learning. In particular, the individual error and the collective error improve over time. Crucially, the effect of social influence on performance is moderated by the selection criterion of who is in the powerful position in the communication network, and by the declaration of the selection criterion. In conclusion, if teams want to utilize the wisdom of crowds within their team, they should admit interaction and opinion exchange, but prevent single individuals from becoming overly influential.

A Appendix: Additional Tables and Figures

	(1) T0 random: answer_6	(2) T1 accuracy: answer_6	(3) T2 confidence: answer_6
own weight (pendant)	0.567***	0.405***	0.392***
	(13.70)	(9.35)	(7.87)
center's weight	0.267***	0.449***	0.469***
	(8.82)	(12.72)	(14.13)
other pendants' weight	0.166***	0.146***	0.139**
	(5.95)	(3.78)	(3.14)
N	528	264	264

t statistics in parentheses
robust s.e. clustered subjects
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A.1: Influence weights on pendants' final answer, separately estimated for each treatment. Regression of the pendant's final answer (period 6) on the initial answers (period 1). Coefficients forced to sum up to one.

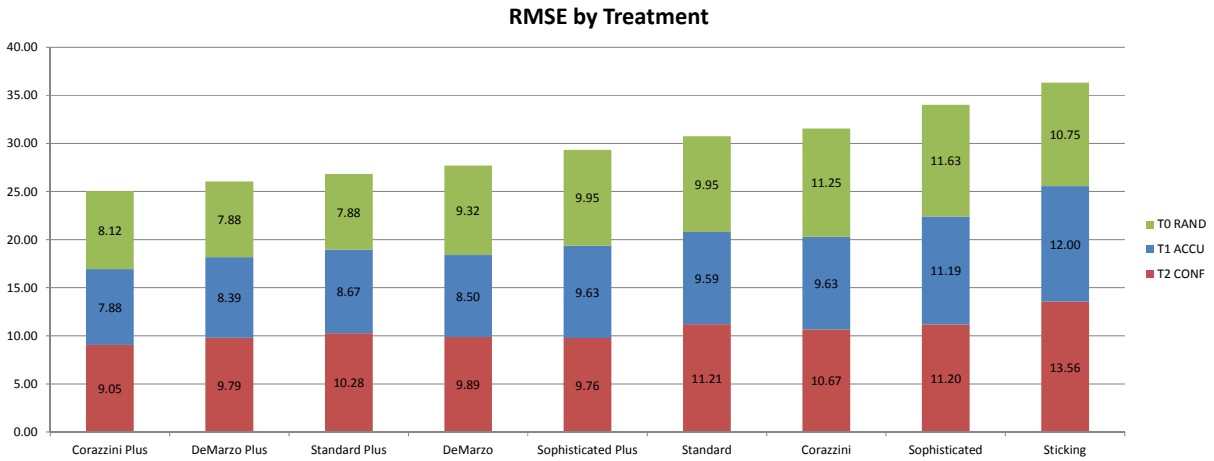


Figure A.1: Root mean squared errors (RMSE) of social learning models differentiated by treatment. Lower errors mean better fit between model and data.

	(1) T0 random: answer_6	(2) T1 accuracy: answer_6	(3) T2 confidence: answer_6
own weight (center)	0.473***	0.659***	0.705***
	(8.80)	(9.91)	(10.68)
pendants' weight	0.527***	0.341***	0.295***
	(9.79)	(5.13)	(4.47)
N	176	88	88

t statistics in parentheses

robust s.e. clustered for 176 subjects

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A.2: Influence weights on center's final answer, separately estimated for each treatment. Regression of the center's final answer (period 6) on the initial answers (period 1). Coefficients forced to sum up to one.

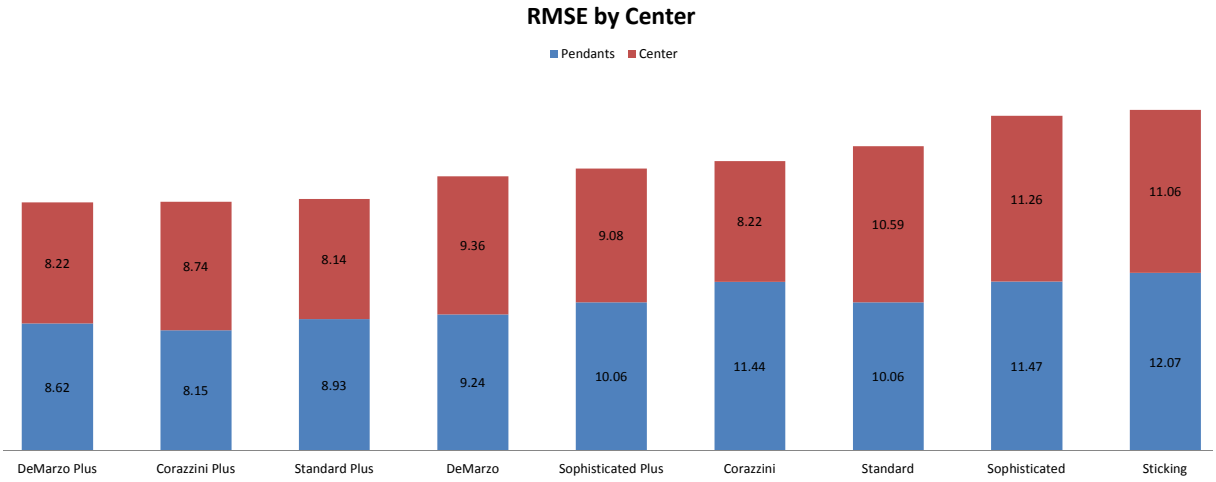


Figure A.2: Root mean squared errors (RMSE) of different models by center and pendants differentiated by center and pendants. Lower errors mean better fit between model and data.

	(1) pendant's answer_6 (last period)
own weight (pendant)	0.577***
	(12.05)
center weight	0.244***
	(6.81)
other pendants weight	0.198***
	(5.31)
accuracy-trait × own	-0.0234
	(-0.41)
accuracy-trait × center	0.0693
	(1.69)
accuracy-trait × other pendants	-0.0393
	(-0.90)
accuracy-declaration (T1) × own	-0.140
	(-1.90)
accuracy-declaration (T1) × center	0.120*
	(2.33)
accuracy-declaration (T1) × other pendants	0.0222
	(0.40)
confidence-trait × own	-0.00712
	(-0.14)
confidence-trait × center	0.0317
	(0.79)
confidence-trait × other pendants	-0.0516
	(-1.15)
confidence-declaration (T2) × own	-0.152*
	(-2.37)
confidence-declaration (T2) × center	0.169***
	(3.39)
confidence-declaration (T2) × other pendants	0.0407
	(0.80)
N	1.056

t statistics in parentheses

robust s.e. clustered for 176 subjects

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A.3: Influence weights on pendant's final answer. Linear regression of the pendant's final answer (period 6) on the initial answers (period 1).

	(1)
	center's answer_6 (last period)
own weight (center)	0.400***
	(6.23)
pendants weight	0.643***
	(9.76)
accuracy-trait × own	0.158*
	(2.45)
accuracy-trait pendants	-0.147*
	(-2.09)
accuracy-declaration (T1) × own	0.0402
	(0.44)
accuracy-declaration (T1) × pendants	-0.0393
	(-0.38)
confidence-trait × own	0.139*
	(2.05)
confidence-trait × pendants	-0.189**
	(-2.70)
confidence-declaration (T2) × own	0.108
	(1.44)
confidence-declaration (T2) × pendants	-0.0353
	(-0.42)
N	352

t statistics in parentheses

robust s.e. clustered for 176 subjects

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A.4: Influence weights on center's final answer. Linear regression of the center's final answer (period 6) on the initial answers (period 1)

	(1) individual error (log)	(2) collective error (log)	(3) crowd error
accuracy-trait	-0.0989	-0.0589	0.0143
	(-1.70)	(-0.62)	(0.05)
accuracy-declaration (T1)	0.108	0.0709	0.185
	(1.56)	(0.56)	(0.55)
confidence-trait	-0.136*	-0.264*	-0.695*
	(-2.23)	(-2.37)	(-2.38)
confidence-declaration (T2)	0.238***	0.355*	1.215**
	(3.36)	(2.65)	(2.93)
overprecision center	0.0426**	0.0453*	0.216***
	(2.76)	(2.05)	(3.32)
overprecision pendants (sum)	0.0268**	0.0268	0.102
	(3.04)	(1.83)	(1.92)
intercept	1.696***	1.706***	
	(10.84)	(7.43)	
intercept cut 1			-0.637
			(-0.70)
intercept cut 1			1.154
			(1.33)
N	1.408	352	352

t statistics in parentheses

Question dummy coefficients not shown

Individual error: robust s.e. clustered for 176 subjects

Collective and crowd errors: robust s.e. clustered for 44 groups

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A.5: Treatment effects on final errors: log error, log collective error and wisdom of crowd error (in period 6). Linear regression (models 1 and 2) and ordered logit regression (model 3).

References

Acemoglu, Daron, Kostas Bimpikis, and Asuman Ozdaglar. 2014. "Dynamics of information exchange in endogenous social networks." *Theoretical Economics* 9:41–97.

Acemoglu, Daron, Asuman Ozdaglar, and Ali ParandehGheibi. 2010. "Spread of (mis)information in social networks." *Games and Economic Behavior* 70:194 – 227.

Ambuehl, Sandro and Shengwu Li. 2014. "Belief Updating and the Demand for Information." *Available at SSRN 2461904* .

Aumann, Robert J. 1976. "An elementary proof that integration preserves uppersemicontinuity." *Journal of Mathematical Economics* 3:15–18.

Battiston, Pietro and Luca Stanca. 2014. "Boundedly Rational Opinion Dynamics in Directed Social Networks: Theory and Experimental Evidence." Working Papers 267, University of Milano-Bicocca, Department of Economics.

Bock, Olaf, Ingmar Baetge, and Andreas Nicklisch. 2014. "hroot: Hamburg registration and organization online tool." *European Economic Review* 71:117–120.

Bolton, Patrick, Markus K Brunnermeier, and Laura Veldkamp. 2013. "Leadership, coordination, and corporate culture." *The Review of Economic Studies* 80:512–537.

Çelen, Boğaçhan and Kyle Hyndman. 2012. "Social Learning Through Endogenous Information Acquisition: An Experiment." *Management Science* 58:1525–1548.

Çelen, Boğaçhan and Shachar Kariv. 2005. "An experimental test of observational learning under imperfect information." *Economic Theory* 26:677–699.

Çelen, Boğaçhan, Shachar Kariv, and Andrew Schotter. 2010. "An experimental test of advice and social learning." *Management Science* 56:1687–1701.

Chandrasekhar, Arun G., Horacio Larreguy, and Juan Pablo Xandri. 2016. "Testing Models of Social Learning on Networks: Evidence from a Lab Experiment in the Field." *mimeo* .

Choi, Syngjoo, Douglas Gale, and Shachar Kariv. 2005. "Behavioral aspects of learning in social networks: an experimental study." *Advances in Applied Microeconomics* 13:25–61.

Corazzini, Luca, Filippo Pavesi, Beatrice Petrovich, and Luca Stanca. 2012. "Influential listeners: An experiment on persuasion bias in social networks." *European Economic Review* 56:1276–1288.

DeGroot, Morris H. 1974. "Reaching a Consensus." *Journal of the American Statistical Association* 69:118–121.

DeMarzo, Peter M., Dimitri Vayanos, and Jeffrey Zwiebel. 2003. "Persuasion Bias, Social Influence, And Unidimensional Opinions." *The Quarterly Journal of Economics* 118:909–968.

Fischbacher, Urs. 2007. "z-Tree: Zurich toolbox for ready-made economic experiments." *Experimental Economics* 10:171–178.

Frey, Bruno S. and Margit Osterloh. 2016. "Aleatoric Democracy." Technical report, CESifo Group Munich.

Friedkin, Noah E. 1991. "Theoretical Foundations for Centrality Measures." *The American Journal of Sociology* 96:1478–1504.

Friedkin, Noah E. and Eugene C. Johnsen. 1990. "Social influence and opinions." *Journal of Mathematical Sociology* 15:193–206.

Gale, Douglas and Shachar Kariv. 2003. "Bayesian learning in social networks." *Games and Economic Behavior* 45:329–346.

Gervais, Simon and Itay Goldstein. 2007. "The positive effects of biased self-perceptions in firms." *Review of Finance* 11:453–496.

Golub, Benjamin and Matthew O. Jackson. 2010. "Naïve Learning in Social Networks and the Wisdom of Crowds." *American Economic Journal: Microeconomics* 2:112–49.

Golub, Benjamin and Matthew O Jackson. 2012. "How homophily affects the speed of learning and best-response dynamics." *The Quarterly Journal of Economics* 127:1287–1338.

Grimm, Veronika and Friederike Mengel. 2016. "An Experiment on Belief Formation in Networks." *Available at SSRN 2361007* .

Haslam, S. Alexander, Craig McGarty, Patricia M. Brown, Rachael A. Eggins, Brenda E. Morrison, and Katherine J. Reynolds. 1998. "Inspecting the emperor's clothes: Evidence that random selection of leaders can enhance group performance." *Group Dynamics: Theory, Research, and Practice* 2:168–184.

Herz, Holger, Daniel Schunk, and Christian Zehnder. 2014. "How do judgmental overconfidence and overoptimism shape innovative activity?" *Games and Economic Behavior* 83:1–23.

Keuschnigg, Marc and Christian Ganser. 2017. "Crowd Wisdom Relies on Agents' Ability in Small Groups with a Voting Aggregation Rule." *Management Science* 63:818–828.

Lorenz, Jan, Heiko Rauhut, Frank Schweitzer, and Dirk Helbing. 2011. "How social influence can undermine the wisdom of crowd effect." *Proceedings of the National Academy of Sciences* 108:9020–9025.

Mannes, Albert E. 2009. "Are We Wise About the Wisdom of Crowds? The Use of Group Judgments in Belief Revision." *Management Science* 55:1267–1279.

Mannes, Albert E. and Don A. Moore. 2013. "A Behavioral Demonstration of Overconfidence in Judgment." *Psychological Science* 24:1190–1197.

Mobius, Markus M., Muriel Niederle, Paul Niehaus, and Tanya S. Rosenblat. 2011. "Managing self-confidence: Theory and experimental evidence." Technical report, National Bureau of Economic Research.

Moore, Don A. and Paul J. Healy. 2008. "The trouble with overconfidence." *Psychological Review* 115:502–517.

Moussaïd, Mehdi, Juliane E. Kämmer, Pantelis P. Analytis, and Hansjörg Neth. 2013. "Social Influence and the Collective Dynamics of Opinion Formation." *PLOS ONE* 8:1–8.

Mueller-Frank, Manuel. 2013. "A general framework for rational learning in social networks." *Theoretical Economics* 8:1–40.

Peterson, Cameron R. and Lee R. Beach. 1967. "Man as an intuitive statistician." *Psychological Bulletin* 68:29.

Phan, Tuan, Adam Szeidl, and Markus Mobius. 2015. "Treasure Hunt: A Field Experiment on Social Learning." mimeo, Society for Economic Dynamics.

Rauhut, Heiko and Jan Lorenz. 2011. "The wisdom of crowds in one mind: How individuals can simulate the knowledge of diverse societies to reach better decisions." *Journal of Mathematical Psychology* 55:191–197.

Rosenberg, Dinah, Eilon Solan, and Nicolas Vieille. 2009. "Informational externalities and emergence of consensus." *Games and Economic Behavior* 66:979–994.

Soll, Jack B. and Joshua Klayman. 2004. "Overconfidence in interval estimates." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 30:299.

Surowiecki, J. 2004. *The Wisdom of Crowds*. New York: Random House.

Zeitoun, Hossam, Margit Osterloh, and Bruno S. Frey. 2014. "Learning from ancient Athens: Demarchy and corporate governance." *The Academy of Management Perspectives* 28:1–14.

Supplementary Online Material

This supplementary online material belongs to the paper "The Strength of Weak Leaders – An Experiment on Social Influence and Social Learning in Teams" by Berno Buechel, Stefan Klößner, Martin Lochmüller, & Heiko Rauhut. It consists of the following sections:

B Mathematical Appendix

B.1 Appendix for Section 3

B.1.1 Theoretical Framework

The uncertainty is described by a probability space $(\Omega, \mathcal{F}, P)$, with Ω being the set of all states of nature, $\mathcal{F}$ being the σ-algebra of events, and P being a probability measure on $\mathcal{F}$. For state of nature $\omega \in \Omega$, the correct answer to the question is denoted by $\theta(\omega)$, i.e., θ is a random variable on $(\Omega, \mathcal{F}, P)$. When the team members are confronted with the question, every team member i is equipped with some information set describing i's knowledge about the true state of nature, $\mathcal{F}_i(0)$, with $i = 1, 2, 3, 4$ denoting the four team members. Thereby, $\mathcal{F}_i(0)$, technically a sub-σ-algebra of $\mathcal{F}$, contains all those events of which team member i knows at time $t = 0$ for sure whether they have occurred or not.

Building only on the information available to them at time $t = 0$, all team members then state their guesses on the correct answer: we denote these answers at time $t = 1$ by $X_i(1)$ $(i = 1, 2, 3, 4)$: the fact that team member i can only make use of the information contained in $\mathcal{F}_i(0)$ technically translates into $X_i(1)$ being a random variable which must be $\mathcal{F}_i(0)$-measurable. Additionally, at time $t = 1$, team member i also provides information about the confidence level associated with $X_i(1)$: this confidence statement will be denoted by $C_i(1)$, technically it is also a $\mathcal{F}_i(0)$-measurable random variable.

After the team members have stated their answers and confidence levels at time $t = 1$, the team leader learns about the other team members' answers, $X_2(1)$, $X_3(1)$, and

$X_4(1)$, as well as their confidence levels, $C_2(1)$, $C_3(1)$, and $C_4(1)$.[1] Thus, the team leader can update by combining the initial information, $\mathcal{F}_1(0)$, and the observed answers and confidence levels of the other team members to build

$$\mathcal{F}_1(1) := \sigma\left(\mathcal{F}_1(0), X_2(1), X_3(1), X_4(1), C_2(1), C_3(1), C_4(1)\right).^2$$

Similarly, the non-central team members can update their information, however, they only observe the answer and confidence level stated by the team leader:

$$\mathcal{F}_i(1) := \sigma\left(\mathcal{F}_i(0), X_1(1), C_1(1)\right) \ , \ i = 2, 3, 4.$$

Again, all team members i now state their answers, $X_i(2)$, and confidence levels, $C_i(2)$. When stating these, team members can only build on the information set $\mathcal{F}_i(1)$, which however in general is larger than $\mathcal{F}_i(0)$, thus the answers and confidence levels stated at time $t = 2$ may well differ from those stated at time $t = 1$.

After the answers and confidence levels at time $t = 2$ have been stated, the team leader again observes what the other team members have stated, which can be used for updating information:

$$\begin{aligned}\mathcal{F}_1(2) &:= \sigma\left(\mathcal{F}_1(1), X_2(2), X_3(2), X_4(2), C_2(2), C_3(2), C_4(2)\right) \\ &= \sigma((\mathcal{F}_1(0), X_i(\tau), C_i(\tau), i = 2, 3, 4, \tau = 1, 2)\,.\end{aligned}$$

Similarly, the non-central team members can update their information, using the team leader's stated answer and confidence level:

$$\mathcal{F}_i(2) := \sigma\left(\mathcal{F}_i(1), X_1(2), C_1(1)\right) = \sigma\left(\mathcal{F}_i(0), X_1(\tau), C_1(\tau), \tau = 1, 2\right) \ , \ i = 2, 3, 4.$$

Yet again, all team members i now state their answers, $X_i(3)$, and confidence levels, $C_i(3)$. When stating these, team members can only build on the information set $\mathcal{F}_i(2)$, which however in general is larger than $\mathcal{F}_i(1)$, thus the answers and confidence levels stated at time $t = 3$ may differ from those stated at time $t = 2$. Afterwards, information updating takes place again, and the process of updating and stating answers and confidence levels goes on. Formally, this can described by $X_i(t)$ and $C_i(t)$ being $\mathcal{F}_i(t-1)$-measurable for all team members $i = 1, 2, 3, 4$ and all times $t = 1, \ldots, 6$, and

$$\begin{aligned}\mathcal{F}_1(t) &:= \sigma\left(\mathcal{F}_1(t-1), X_2(t), X_3(t), X_4(t), C_2(t), C_3(t), C_4(t)\right) \\ &= \sigma((\mathcal{F}_1(0), X_i(\tau), C_i(\tau), i = 2, 3, 4, \tau = 1, \ldots, t)\end{aligned}$$

[1] In this mathematical appendix we use capital letters to indicate random variables.

[2] $\sigma(\cdot)$ denotes the result of combining information, technically, it is the smallest sub-σ-algebra of $\mathcal{F}$ with respect to which all combined information is measurable.

as well as

$$\mathcal{F}_i(t) := \sigma\left(\mathcal{F}_i(t-1), X_1(t), C_1(t)\right) = \sigma\left(\mathcal{F}_i(0), X_1(\tau), C_1(\tau), \tau = 1, \ldots, t\right) \ , \ i = 2, 3, 4$$

for all times $t = 1, \ldots, 6$.

Using a payoff function, Π, which is decreasing in its argument, team member i's guess at time t, $X_i(t)$, is awarded by $\Pi(|\theta - X_i(t)|)$. In the end, the actual payoff is determined by randomly choosing the payoff belonging to one of the six answers, i.e., the payoff equals $\Pi(|\theta - X_i(1)|), \ldots, \Pi(|\theta - X_i(6)|)$, each with a probability of 1/6.

B.1.2 Rational Models of Learning

Rational approaches assume that team members maximize their expected payoff. According to rational models, team member i will choose $X_i(1), \ldots, X_i(6)$ and $C_i(1), \ldots, C_i(6)$ such that the expected payoff

$$\frac{1}{6} \sum_{t=1}^{6} E\left(\Pi(|\theta - X_i(t)|)\right)$$

becomes as large as possible.

First, we state an almost trivial lemma about the maximal amount of information the team members can collect.

Lemma B.1. *Information acquisition in the team is bounded, no team member can learn more than the combination of all team members' initial information, technically:*

$$\mathcal{F}_i(t) \subseteq \sigma\left(\mathcal{F}_1(0), \mathcal{F}_2(0), \mathcal{F}_3(0), \mathcal{F}_4(0)\right) =: \mathcal{F}(0).$$

We now discuss how the team leader is expected to behave under rational models of learning.

Proposition B.1. *1. If the information contained in the pendants' first-round answers and confidence statements allows the team leader to get to know all of the information contained in the pendants' initial information that is important with respect to the correct answer, then the team leader will give the same, optimal answer in rounds 2 through 6. Formally,*

$$\textit{if } P\left(\theta | \sigma(\mathcal{F}_1(0), X_i(1), C_i(1), i = 2, 3, 4)\right) = P\left(\theta | \mathcal{F}(0)\right),$$

$$\textit{then } X_1(t) = \underset{X \mathcal{F}(0) - \textit{measurable}}{\arg\max} E\left(\Pi\left(|\theta - X|\right)\right) \textit{ for } t = 2, \ldots, 6.$$

This is in particular fulfilled if the team leader is able to completely infer the maximally available information, $\mathcal{F}(0)$, from the other team members' first round answers and confidence statements, i.e., if $\sigma(\mathcal{F}_1(0), X_i(1), C_i(1), i = 2, 3, 4) = \mathcal{F}(0)$.

2. *If $P(\theta|\sigma(\mathcal{F}_1(0), X_i(1), C_i(1), i = 2, 3, 4)) = P(\theta|\mathcal{F}(0))$ (as in '1.'), then the team leader's optimal behavior is to give the answers $X^* := \underset{X\mathcal{F}(0)-measurable}{\arg\max} E(\Pi(|\theta - X|))$ in rounds $t = 2, \ldots, 6$ and $\underset{X\mathcal{F}_1(0)-measurable}{\arg\max} E(\Pi(|\theta - X|))$ in the first round.*

Proof. 1. Because of Lemma B.1, the team leader can never give an answer better than

$$\underset{X\mathcal{F}(0)-\text{measurable}}{\arg\max} E(\Pi(|\theta - X|)).$$

On the other hand, given that

$$P(\theta|\sigma(\mathcal{F}_1(0), X_i(1), C_i(1), i = 2, 3, 4)) = P(\theta|\mathcal{F}(0)),$$

the team leader can form this conditional expectation at times $t = 2, \ldots, 6$, because it can be formed when knowing $\mathcal{F}_1(0)$, $X_2(1)$, $X_3(1)$, $X_4(1)$, $C_2(1)$, $C_3(1)$, and $C_4(1)$.

2. The statement for rounds 2 through 6 has already been proven in '1.', and the statement for the first round follows from the same reasons. As this strategy separately maximizes each of the terms in the expected payoff, $\frac{1}{6}\sum_{t=1}^{6} E(\Pi(|\theta - X_1(t)|))$, it is the optimal strategy for the team leader.

□

We now discuss how the pendants are expected to behave under rational models of learning.

Proposition B.2. *1. If, from the team leader's answers and confidence statements in the first two rounds, pendant i can learn everything that is relevant with respect to the correct answer, then pendant i will state the optimal answer in rounds 3 through 6. Formally,*

$$\text{if } P(\theta|\sigma(\mathcal{F}_i(0), X_1(1), C_1(1), X_1(2), C_1(2))) = P(\theta|\mathcal{F}(0)),$$

$$\text{then } X_i(t) = \underset{X\mathcal{F}(0)-measurable}{\arg\max} E(\Pi(|\theta - X|)) \text{ for } t = 3, \ldots, 6.$$

This is in particular fulfilled if pendant i is able to completely infer the maximally available information, $\mathcal{F}(0)$, from the team leader's first and second round answers and confidence statements, i.e., if $\sigma(\mathcal{F}_i(0), X_1(1), C_1(1), X_1(2), C_1(2)) = \mathcal{F}(0)$.

2. *If* $P(\theta|\sigma(\mathcal{F}_i(0), X_1(1), C_1(1), X_1(2), C_1(2))) = P(\theta|\mathcal{F}(0))$ *(as in '1.'), then pendant i's optimal strategy is to give the answers* $\underset{X\mathcal{F}(0)-measurable}{\arg\max} E(\Pi(|\theta - X|))$ *in rounds* $t = 3, \ldots, 6$, $\underset{X\mathcal{F}_i(1)-measurable}{\arg\max} E(\Pi(|\theta - X|))$ *in the second round, as well as* $\underset{X\mathcal{F}_i(0)-measurable}{\arg\max} E(\Pi(|\theta - X|))$ *in the first round.*

Proof. The proofs are analogous to the corresponding proofs of Proposition B.1. □

Overall, we have thus derived the following results which correspond to Prediction 1: if answers and confidence statements of the team members can be used to gain all relevant information contained in the team members' initial information, then the team leader will state the optimal answer in rounds 2 through 6 and the pendants will state the optimal answer in rounds 3 through 6.

B.1.3 Naïve Models of Learning

Naive models of learning suppose that, from round to round, answers are convex combinations of own and other team members' answers according to weights g_{ij}:

$$X_1(t+1) = g_{11}X_1(t) + g_{12}X_2(t) + g_{13}X_3(t) + g_{14}X_4(t),$$
$$X_i(t+1) = g_{i1}X_1(t) + g_{ii}X_i(t), i = 2, 3, 4.$$

Using the notation $X(t) := (X_1(t), \ldots, X_4(t))'$ for $t = 1, \ldots, 6$, the updating can conveniently be written in vector and matrix notation as $X(t+1) = GX(t)$, where G is given as follows:

$$G = \begin{pmatrix} g_{11} & g_{12} & g_{13} & g_{14} \\ g_{21} & g_{22} & 0 & 0 \\ g_{31} & 0 & g_{33} & 0 \\ g_{41} & 0 & 0 & g_{44} \end{pmatrix}. \tag{B.1}$$

G is a row-stochastic matrix which means that all entries of G are non-negative and that, for each row, the sum of the corresponding entries equals unity. Additionally, to avoid trivial special cases, we assume that all the parameters in equation (B.1) are strictly positive: $g_{11}, g_{1i}, g_{i1}, g_{ii} > 0$ for $i = 2, 3, 4$, meaning that, when updating, the team leader takes into account the previous guesses of all team members, while all other team members update their guesses using their own and the team leader's previous guess.[3]

We first discuss under which conditions the team leader and pendants update their guesses only once and twice, respectively.

[3]In section 5.1, we study one baseline model, called the *Sticking Model*, in which this assumption is not satisfied. In that model, we have $g_{ii} = 1$ for all $i = 1, ..., 4$ and hence $g_{ij} = 0$ for $i \neq j$.

Proposition B.3. *1. The team leader's guess is updated only once if and only if all team members put identical weights to the team leader when updating. Formally:* $(1,0,0,0)'G = (1,0,0,0)'G^t$ *for all* t *if and only if* $g_{11} = g_{21} = g_{31} = g_{41}$.

2. If the team leader's guess is updated only once, then the other team members update their guesses more than twice.

Proof. 1. First of all, notice that the team leader's guess at time t is given by

$$(1,0,0,0)'G^{t-1}X(1).$$

Furthermore, if $(1,0,0,0)'G = (1,0,0,0)'G^2$, then $(1,0,0,0)'G^{t-1} = (1,0,0,0)'G$ for all t. We therefore only have to consider the first rows of G and G^2. For $i = 2,3,4$, the i-th element of the first row of G^2 is easily seen to be $g_{1i}(g_{11} + g_{ii})$. It equals the corresponding element of G if and only if the equation $g_{1i} = g_{1i}(g_{11} + g_{ii})$ holds. This is equivalent to $g_{11} + g_{ii} = 1$ for $i = 2,3,4$, which in turn is equivalent to $g_{1i} = 1 - g_{ii} = g_{11}$ for $i = 2,3,4$, because G is row-stochastic.

2. Similar to above, for checking whether the team members update more than twice, it suffices to check whether the corresponding rows of G^2 equal those of G^3. We exemplarily consider the second team member and calculate $(0,1,0,0)'G^2$ as well as $(0,1,0,0)'G^3$, the calculations for $i = 3,4$ are completely analogous. When the team leader updates only once, the row-stochastic matrix G can be written as follows:

$$G = \begin{pmatrix} g_{11} & g_{12} & g_{13} & g_{14} \\ g_{11} & 1-g_{11} & 0 & 0 \\ g_{11} & 0 & 1-g_{11} & 0 \\ g_{11} & 0 & 0 & 1-g_{11} \end{pmatrix},$$

with $g_{11} = 1 - g_{12} - g_{13} - g_{14}$. From this, we find:

$$(0,1,0,0)'\,G^2 = \left(g_{11}, g_{12}g_{11} + (1-g_{11})^2, g_{13}g_{11}, g_{14}g_{11}\right)',$$
$$(0,1,0,0)'\,G^3 = \left(g_{11}, g_{12}g_{11}(2-g_{11}) + (1-g_{11})^3, g_{13}g_{11}(2-g_{11}), g_{14}g_{11}(2-g_{11})\right)'.$$

Thus, the second team member will in general state different guesses after the second and third updating.

□

Hence, we have a clear difference to the rational models. According to the naïve models, the team leader will update more than once, except for the special case that all pendants put the same weight on the team leader's previous guess when updating. If

this special case is given, the pendants will update more than twice. Hence, under naïve models, it is not possible that agents state an optimal answer from round $t = 3$ on.

We now turn our attention to the quality of learning, by studying the mean absolute error of the team members' guesses at the beginning and after the first round of communication.

Lemma B.2. *Let $Y_1, \ldots, Y_n$ be n random variables and denote the corresponding mean absolute errors by $\mathrm{MAE}_{Y_i} := E(|Y_i - \theta|)$ for $i = 1, \ldots, n$. Let further $Y := \lambda_1 Y_1 + \ldots + \lambda_n Y_n$ be a convex combination of $Y_1, \ldots, Y_n$ with non-negative weights λ summing to unity and denote the corresponding mean absolute error by $\mathrm{MAE}_Y := E(|Y - \theta|)$. Then we have:*

$$\mathrm{MAE}_Y \leq \lambda_1 \mathrm{MAE}_{Y_1} + \ldots + \lambda_n \mathrm{MAE}_{Y_n}, \tag{B.2}$$

with equality if and only if $P\left((Y_1 \geq \theta \wedge \ldots \wedge Y_n \geq \theta) \vee (Y_1 \leq \theta \wedge \ldots \wedge Y_n \leq \theta)\right) = 1$.

Proof. First of all

$$\begin{aligned} |Y - \theta| &= |\lambda_1 Y_1 + \ldots + \lambda_n Y_n - \theta| \\ &= |\lambda_1 (Y_1 - \theta) + \ldots + \lambda_n (Y_n - \theta)| \\ &\leq \lambda_1 |Y_1 - \theta| + \ldots + \lambda_n |Y_n - \theta|, \end{aligned}$$

from which taking expectations yields

$$\mathrm{MAE}_Y \leq \lambda_1 E(|Y_1 - \theta|) + \ldots + \lambda_n E(|Y_n - \theta|) = \lambda_1 \mathrm{MAE}_{Y_1} + \ldots + \lambda_n \mathrm{MAE}_{Y_n},$$

with equality if and only if $|\lambda_1 (Y_1 - \theta) + \ldots + \lambda_n (Y_n - \theta)|$ equals $\lambda_1 |Y_1 - \theta| + \ldots + \lambda_n |Y_n - \theta|$ almost surely. Thus, to complete the proof, we only have to show that the latter happens if and only if $a_1 := Y_1 - \theta, \ldots, a_n := Y_n - \theta$ are either all non-negative or all non-positive almost surely. To this end, we compute $|a_1 + \ldots + a_n|^2 = (a_1 + \ldots + a_n)^2$ and compare this quantity to $(|a_1| + \ldots + |a_n|)^2$. For the first quantity, we find $a_1^2 + \ldots + a_n^2 + \sum_{i \neq j} a_i a_j$, while the second quantity equals $a_1^2 + \ldots + a_n^2 + \sum_{i \neq j} |a_i| \, |a_j|$. The two quantities are thus equal if and only if $a_i a_j = |a_i a_j|$ for all i, j, which only happens if $a_1, \ldots, a_n$ are either all non-negative or all non-positive. □

The following lemma is an immediate consequence of Lemma B.2.

Lemma B.3. *For the weighted average $X_{2:4}(1) := \lambda_2 X_2(1) + \lambda_3 X_3(1) + \lambda_4 X_4(1)$ of the non-leaders' opinions $X_2(1)$, $X_3(1)$, and $X_4(1)$, with $\lambda_i := \frac{g_{1i}}{g_{12} + g_{13} + g_{14}}$ for $i = 2, 3, 4$, we have:*

$$\mathrm{MAE}_{2:4}(1) \leq \lambda_2 \mathrm{MAE}_2(1) + \lambda_3 \mathrm{MAE}_3(1) + \lambda_4 \mathrm{MAE}_4(1), \tag{B.3}$$

where $\mathrm{MAE}_{2:4}(1) := E\left(|X_{2:4}(1) - \theta|\right)$ *and* $\mathrm{MAE}_i(1) := E\left(|X_i(1) - \theta|\right)$ *for* $i = 2, 3, 4$. *In equation* (B.3)*, equality holds if and only if* $X_2(1)$*,* $X_3(1)$*, and* $X_4(1)$ *lie on the same side of* θ *almost surely, i.e., if*

$$P\Big(\big(X_2(1), X_3(1), X_4(1) \geq \theta\big) \vee \big(X_2(1), X_3(1), X_4(1) \leq \theta\big)\Big) = 1.$$

Lemma B.3 shows that averaging the pendants' initial guesses typically is an improvement over their individual initial guesses.

Proposition B.4. *1. If* $\mathrm{MAE}_{2:4}(1) \leq \mathrm{MAE}_1(1)$*, then* $\mathrm{MAE}_1(2) \leq \mathrm{MAE}_1(1)$*, with equality if and only if* $P\big((X_1(1) \geq \theta \wedge X_{2:4}(1) \geq \theta) \vee (X_1(1) \leq \theta \wedge X_{2:4}(1) \leq \theta)\big) = 1$.

2. For $i = 2, 3, 4$*: if* $\mathrm{MAE}_1(1) \leq \mathrm{MAE}_i(1)$*, then* $\mathrm{MAE}_i(2) \leq \mathrm{MAE}_i(1)$.

Proof. 1. Since $X_1(2) = g_{11}X_1(1) + (g_{12} + g_{13} + g_{14})X_{2:4}(1)$ with $g_{11} + g_{12} + g_{13} + g_{14} = 1$, Lemma B.2 implies $\mathrm{MAE}_1(2) \leq g_{11}\,\mathrm{MAE}_1(1) + (1 - g_{11})\,\mathrm{MAE}_{2:4}(1)$, from which the assertion follows immediately.

2. Applying Lemma B.2 to $X_i(2) = g_{i1}X_1(1) + g_{ii}X_i(1)$ yields $\mathrm{MAE}_i(2) \leq g_{i1}\,\mathrm{MAE}_1(1) + g_{ii}\,\mathrm{MAE}_i(1)$, from which the assertion follows immediately.

□

Proposition B.4 shows that the team leader's guess will on average improve from the first to the second round if the combination of the other team members' initial guesses is a signal that is not worse than the team leader's initial one. This is a realistic assumption, particularly under the random treatment T0. Furthermore, a pendant's guess will improve after the first updating if the team leader's initial guess is on average not worse than that pendant's signal. This is a realistic assumption, particularly for treatments T1 accuracy and T2 confidence.

B.2 Appendix for Section 5.1

B.2.1 Specific Rational Models

Building on the theoretical framework laid out above, we will now consider specific rational models, by specifying in particular what information team members initially possess. To start, however, we discuss how correct answers are modeled.

For ease of presentation, we interpret the correct answers to the questions asked in our experiment as points in $[0, 1]$, although answers had to be integer numbers between 0 and 100. For instance, 71, the correct answer to the question about the voter turnout to the federal elections in Germany in 2009, is translated into 0.71 and could also be interpreted as the probability of a randomly chosen eligible voter actually casting a ballot. This

given, we assume that the prior, unconditional distribution of the correct answer is the uniform distribution on the unit interval: $\theta \sim U(0,1)$, with probability density function (pdf) $f_\theta(p) = 1$ for all $p \in [0,1]$, meaning that, a priori, before any agent has received any information, all answers were equally likely to be the correct one. The uniform distribution corresponds to a beta distribution $\beta(1,1)$ which was originally suggested as *the* prior distribution by Thomas Bayes. With respect to initial information, we assume that each team member i ($i = 1, \ldots, 4$) observes a two-dimensional signal $\psi_i = (S_i, F_i)$ which is, conditional on θ, stochastically independent from the other team members' signals. This signal can be interpreted in the following way: every team member i has some pool of observations, where observations can either be 'successes' or 'failures', and the number of successes is S_i, while the number of failures is F_i. Here, 'successes' and 'failures' mean that the condition asked for is fulfilled or not: in case of the voter turnout, S_i gives the number of people of which team member i knows that they cast a vote, while F_i denotes the number of people of which team member i knows that they abstained from voting.

The 'Standard' Model

In the 'Standard' model, it is assumed that all team members possess the same amount of information, i.e., that $S_1 + F_1 = \ldots = S_4 + F_4$.[4] With respect to the link between the distribution of (S_i, F_i) to the unknown, correct answer, we assume the following: the probability of observing $(S_i, F_i) = (s_i, f_i)$ is, conditional on $\theta = p$, given by[5] $\binom{s_i+f_i}{s_i} \cdot p^{s_i} \cdot (1-p)^{f_i}$. Put differently, the number of successes follows, conditional on $\theta = p$, a binomial distribution with parameters p and $n := s_i + f_i$. Observing the signal, team member i may update the a priori belief by using Bayes' rule, forming the distribution of θ conditional on observing $S_i = s_i$:

$$f_{\theta|S_i=s_i,F_i=f_i}(p) = \frac{\binom{s_i+f_i}{s_i} \cdot p^{s_i} \cdot (1-p)^{f_i}}{\int_0^1 \binom{s_i+f_i}{s_i} \cdot \widetilde{p}^{s_i} \cdot (1-\widetilde{p})^{f_i} d\widetilde{p}} = \frac{p^{s_i} \cdot (1-p)^{f_i}}{B(s_i+1, f_i+1)},$$

with $B(\alpha, \beta)$ denoting Euler's beta function. Thus, team member i's inital belief before communication is a beta distribution with parameters $s_i + 1$ and $f_i + 1$. Therefore, team

[4]This number of observations may be some fixed integer, n, or, more generally, a random variable N taking integer values.

[5]If the number of observations is a random variable, then one also conditions on $N = n$, and $P(N = n) \cdot$ appears as an additional factor.

member i's optimal answer in the first round is p^*, with p^* maximizing

$$\int_0^1 \Pi(|p-p^*|) f_{\theta|S_i=s_i,F_i=f_i}(p)dp = \int_0^1 \Pi(|p-p^*|) \frac{p^{s_i}\cdot(1-p)^{f_i}}{B(s_i+1,f_i+1)}dp.$$

Due to the specific structure of the payoff function used in our experiment, the beta distribution's mode, $\frac{s_i}{s_i+f_i} = \frac{s_i}{n}$, is a very good approximation to p^*, we will therefore assume that team member i states the answer $X_i(1) = \frac{s_i}{s_i+f_i} = \frac{s_i}{n}$ in the first round. Continuing the example on voter turnout: If an individual knows about ten citizens that seven of them voted and three of them abstained, then his belief is beta distributed with a mode of $\frac{7}{7+3} = 0.7$, which is his initial guess.

After the first round of answers, the team leader gets to know the answers of all team members, thus the team leader can easily recover s_2, s_3, and s_4 to gain the maximally available information, $\mathcal{F}(0)$. The corresponding belief upon maximal information, i.e., upon observing $s_1,\ldots,s_4$, is

$$f_{\theta|S_1=s_1,\ldots,S_4=s_4,F_1=f_1,\ldots,F_4=f_4}(p) = \frac{p^{s_1+\ldots+s_4}\cdot(1-p)^{f_1+\ldots+f_4}}{B(1+\sum_{i=1}^4 s_i, 1+\sum_{i=1}^4 f_i)},$$

again a beta distribution, with parameters $1+\sum_{i=1}^4 s_i$ and $1+\sum_{i=1}^4 f_i$. Thus, the team leader's optimal answer in rounds 2 through 6 is the corresponding mode, $\frac{s_1+\ldots+s_4}{s_1+\ldots+s_4+f_1+\ldots+f_4} = \frac{s_1+\ldots+s_4}{4n}$, which can be rewritten as $\frac{1}{4}X_1(1)+\ldots+\frac{1}{4}X_4(1)$, an equally weighted average of the team members' first-round answers.

The other team members' second-round answers can be built using only the corresponding initial signal, ψ_i, as well as the team leader's first-round answer, $X_1(1)$. The latter allows to infer s_1, thus team member i's knowledge in the second round consists of s_1 and s_i. Analogously to above, it is easy to derive that the corresponding belief is again a beta distribution, with parameters $1+s_1+s_i$ and $1+f_1+f_i$. The corresponding optimal answers in the second round thus are $\frac{1}{2}X_1(1)+\frac{1}{2}X_i(1)$, while from round 3 on, the team leader's optimal answer will be copied.

Overall, the updating in the 'Standard' model can be summarized as follows:

Summary 1 (*Standard Model*)**.** *The team leader computes the unweighted average of all team members' first-round answers and states* $\frac{1}{4}X_1(1)+\ldots+\frac{1}{4}X_4(1)$ *from round* 2 *on, the other team members state* $\frac{1}{2}X_1(1)+\frac{1}{2}X_i(1)$ *in round* 2*, and they join the team leader in stating the average of the team's first-round answers from round* 3 *on.*

The 'Sophisticated' Model

For the 'Sophisticated' model, the link between the distribution of the signal $\psi_i = (S_i, F_i)$ to the unknown, correct answer, θ, looks as follows: the probability of observing $(S_i, F_i) = (s_i, f_i)$ is, conditional on $\theta = p$ and $S_i + F_i = n_i := s_i + f_i$, given by $\binom{n}{s_i} \cdot p^{s_i} \cdot (1-p)^{f_i}$. Put differently, the number of successes follows, conditional on $\theta = p$ and $S_i + F_i = n_i$, a binomial distribution with parameters p and $n_i = S_i + F_i$. As for the 'Standard' model, one easily derives that team member i's belief about the correct answer is a beta distribution with parameters $1 + s_i$ and $1 + f_i$, implying that the first-round answer is $\frac{s_i}{s_i+f_i} = \frac{s_i}{n_i}$.

In our experiment, team members were not only asked about their guess with respect to the correct answer to the question at hand, but they also supplied a measure of the confidence in their answer. More precisely, they essentially provided an interval that should contain the correct answer with a probability of 90%. Based on the Beta$(1+s_i, 1+f_i)$-belief, team member i's first-round statement thus does not only consist of guessing the correct answer by $X_i(1) = \frac{s_i}{n_i}$, but also of supplying the corresponding confidence $C_i(1)$ which is a function of s_i and f_i, $C_i(1) = \text{Conf}(s_i, f_i)$.

After the first round of answers, the team leader gets to know not only the answers of all team members, $X_i(1) = \frac{s_i}{s_i+f_i}$, but also their confidence statements, $C_i(1) = \text{Conf}(s_i, f_i)$ $(i = 2, 3, 4)$. Using these, the team leader can recover s_2, s_3, and s_4 as well as f_2, f_3, and f_4, to gain the maximally available information, $\mathcal{F}(0)$. The corresponding belief upon maximal information, i.e., upon observing $s_1, \ldots, s_4$, $f_1, \ldots, f_4$, is

$$f_{\theta|S_1=s_1,\ldots,S_4=s_4,F_1=f_1,\ldots,F_4=f_4}(p) = \frac{p^{s_1+\ldots+s_4} \cdot (1-p)^{f_1+\ldots+f_4}}{B(1+\sum_{i=1}^{4} s_i, 1+\sum_{i=1}^{4} f_i)},$$

again a beta distribution, with parameters $1+\sum_{i=1}^{4} s_i$ and $1+\sum_{i=1}^{4} f_i$. Thus, the team leader's optimal answer in rounds 2 through 6 is the corresponding mode, $\frac{s_1+\ldots+s_4}{s_1+\ldots+s_4+f_1+\ldots+f_4} = \frac{s_1+\ldots+s_4}{n_1+\ldots+n_4}$, which can be rewritten as $\frac{n_1}{n_1+\ldots+n_4}X_1(1)+\ldots+\frac{n_4}{n_1+\ldots+n_4}X_4(1)$, a weighted average of the team members' first-round answers.

The other team members' second-round answers can be build using only the corresponding initial signal, ψ_i, as well as the team leader's first-round answer and confidence statement, $X_1(1)$ and $C_1(1)$. The latter quantities allow to infer s_1 and f_1, thus team member i's knowledge in the second round consists of s_1, s_i and f_1, f_i. Analogously to above, it is easy to derive that the corresponding belief is again a beta distribution, with parameters $1 + s_1 + s_i$ and $1 + f_1 + f_i$. The corresponding optimal answers in the second round thus are $\frac{s_1+s_i}{n_1+n_i} = \frac{n_1}{n_1+n_i}X_1(1) + \frac{n_i}{n_1+n_i}X_i(1)$, while from round 3 on, the team leader's optimal answer will be copied.

Overall, the updating in the 'Sophisticated' model can be summarized as follows:

Summary 2 (*Sophisticated Model*)**.** *The team leader computes a weighted average of all team members' first-round answers and states*

$$\frac{N_1}{N_1+N_2+N_3+N_4}X_1(1)+\ldots+\frac{N_4}{N_1+N_2+N_3+N_4}X_4(1)$$

from round 2 *on, the other team members state* $\frac{N_1}{N_1+N_i}X_1(1)+\frac{N_i}{N_1+N_i}X_i(1)$ *in round* 2*, and they join the team leader in stating the weighted average of the team's first-round answers from round* 3 *on.*

B.2.2 Models of Rational Learning with Conservatism

In the following, we will enrich the models of rational learning by conservatism that results from overprecision. Overprecision is the empirically observed phenomenon that people typically provide too narrow confidence intervals when asked about their confidence. To model overprecision, we will assume that agents treat their initial private signal as more precise than it actually is. We will further assume that agents account for the fact that other team members are overprecise, but are blind with respect to their own level of overprecision. While the level of overprecision is in principle agent-specific, our analysis focuses on the case in which all agents are equally overprecise. The models with overprecision nest the rational models when setting the level of overprecision to zero.

The 'Standard Plus' model

By the 'Standard Plus' model, we denote the extension of the 'Standard' model by overprecision. The only difference to the 'Standard' model is that we assume that team members misinterpret their signal: when the signal actually is $\psi_i = (s_i, f_i)$, team member i will interpret it as if the received signal was $(\tau s_i, \tau f_i)$, where $\tau \geq 1$ is a parameter to capture overprecision.[6] Therefore, team member i's belief in the first round will be given by a Beta$(1+\tau s_i, 1+\tau f_i)$ distribution, leading to the first-round answer $\frac{\tau s_i}{\tau s_i+\tau f_i} = \frac{s_i}{n}$, as in the 'Standard' model. However, after learning from the other team members, the second-round answers are still prone to overprecision: from round 2 on, the team leader will state $\frac{\tau s_1+s_2+\ldots+s_4}{\tau s_1+s_2+\ldots+s_4+\tau f_1+f_2+\ldots+f_4}$, which can be rewritten as $\frac{\tau}{\tau+3}X_1(1)+\frac{1}{\tau+3}X_2(1)+\frac{1}{\tau+3}X_3(1)+\frac{1}{\tau+3}X_4(1)$. Similarly, other team members will state $\frac{1}{\tau+1}X_1(1)+\frac{\tau}{\tau+1}X_i(1)$ in the second round ($i=2,3,4$). In rounds 3 though 6, however, in contrast to the 'Standard' model, the other team members will not copy the team leader's second-round answer. Instead, driven by overconfidence, team member i will state $\frac{\tau s_i+\sum_{j\neq i} s_j}{\tau s_i+\sum_{j\neq i} s_j+\tau f_i+\sum_{j\neq i} f_j}$, which can be rewritten as $\frac{\tau}{\tau+3}X_i(1)+\sum_{j\neq i}\frac{1}{\tau+3}X_j(1)$.

[6] In our empirical application, τ is fixed to $\tau = 5$, in order to appropriately account for the overprecision inherent in the confidence intervals given by the team members.

Overall, the updating in the 'Standard Plus' model can be summarized as follows:

Summary 3 (*Standard-Plus Model*)**.** *The team leader computes a weighted average of all team members' first-round answers and states* $\frac{\tau}{\tau+3}X_1(1)+\frac{1}{\tau+3}X_2(1)+\ldots+\frac{1}{\tau+3}X_4(1)$ *from round* 2 *on, other team member* i *states* $\frac{1}{\tau+1}X_1(1)+\frac{\tau}{\tau+1}X_i(1)$ *in round* 2 *(*$i=2,3,4$*), while stating* $\frac{\tau}{\tau+3}X_i(1)+\sum_{j\neq i}\frac{1}{\tau+3}X_j(1)$ *from round* 3 *on.*

The 'Sophisticated Plus' model

By the 'Sophisticated Plus' model, we denote the extension of the 'Sophisticated' model by overprecision. The only difference to the 'Sophisticated' model is that we assume that team members misinterpret their signal: as above, when the signal actually is $\psi_i=(s_i,f_i)$, team member i will interpret it as if the received signal was $(\tau s_i,\tau f_i)$. Therefore, team member i's belief in the first round will be given by a Beta$(1+\tau s_i,1+\tau f_i)$ distribution, leading to the first-round answer $\frac{\tau s_i}{\tau s_i+\tau f_i}=\frac{s_i}{n_i}$, as in the 'Sophisticated' model. However, after learning from the other team members, the second-round answers are still biased by overprecision: from round 2 on, the team leader will state $\frac{\tau s_1+s_2+\ldots+s_4}{\tau s_1+s_2+\ldots+s_4+\tau f_1+f_2+\ldots+f_4}$, which can be rewritten as $\frac{\tau n_1}{\tau n_1+n_2+n_3+n_4}X_1(1)+\frac{n_2}{\tau n_1+n_2+n_3+n_4}X_2(1)+\frac{n_3}{\tau n_1+n_2+n_3+n_4}X_3(1)+\frac{n_4}{\tau n_1+n_2+n_3+n_4}X_4(1)$. Similarly, other team members will state $\frac{n_1}{n_1+\tau n_i}X_1(1)+\frac{\tau n_i}{n_1+\tau n_i}X_i(1)$ in the second round ($i=2,3,4$). In rounds 3 though 6, however, in contrast to the *Sophisticated Model*, the other team members will not copy the team leader's second-round answer. Instead, driven by overconfidence, team member i will state $\frac{\tau s_i+\sum_{j\neq i}s_j}{\tau s_i+\sum_{j\neq i}s_j+\tau f_i+\sum_{j\neq i}f_j}$, which can be rewritten as $\frac{\tau n_i}{\tau n_i+\sum_{j\neq i}n_j}X_i(1)+\sum_{j\neq i}\frac{n_j}{\tau n_i+\sum_{j\neq i}n_j}X_j(1)$.

Overall, the updating in the 'Sophisticated Plus' model can be summarized as follows:

Summary 4 (*Sophisticated-Plus Model*)**.** *The team leader computes a weighted average of all team members' first-round answers and states*

$$\frac{\tau N_1}{\tau N_1+N_2+N_3+N_4}X_1(1)+\sum_{j=2}^{4}\frac{N_j}{\tau N_1+N_2+N_3+N_4}X_j(1)$$

from round 2 *on, other team member* i *states* $\frac{N_1}{N_1+\tau N_i}X_1(1)+\frac{\tau N_i}{N_1+\tau N_i}X_i(1)$ *in round* 2 *(*$i=2,3,4$*), while stating*

$$\frac{\tau N_i}{\tau N_i+\sum_{j\neq i}N_j}X_i(1)+\sum_{j\neq i}\frac{N_j}{\tau N_i+\sum_{j\neq i}N_j}X_j(1)$$

from round 3 *on.*

C Appendix: Details of the Experimental Design

The experiment was run in eleven sessions (which followed after two pilot sessions) in August and September 2013. It was conducted in the experimental laboratory of the Faculty of Economic and Social Sciences at the University of Hamburg, Germany. It was programmed using zTree (Fischbacher, 2007) and organized and recruited with *hroot* (Bock et al., 2014). In total 176 university students with various academic backgrounds participated in the experiment. The participants earned on average EUR 9.50. The norm at the lab was EUR 10. Each session lasted approximately 60 minutes, including instructions, questionnaire and payments.

Subjects were randomly assigned to computer terminals. After all participants were seated, two sets of instructions were handed out, a German version as well as an English translation. The German instructions were read aloud to establish common knowledge. The subjects were then given the possibility to ask questions, which were answered privately. The instructions were left with the participants for reference during the whole experiment. In the instructions it was pointed out that the use of mobile phones, smart phones as well as tablets or similar devices would lead to expulsion from the experiment and exclusion from all payments.[7] There were no data exclusions.

All decisions and the payments at the end of the experiment were made anonymously. The participants were not informed about the identity of any other participant and they were paid privately upon completion of each session. The individual computer terminals were separated by boards and could be partially closed with curtains.

C.1 Experimental Task

The design of the experiment draws upon the studies by Lorenz et al. (2011), Rauhut and Lorenz (2011), and Moussaïd et al. (2013). The subjects were asked to give estimates on factual questions and to state their confidence level. The experiment was based on questions with hard facts, because they admit an unambiguously correct answer. For instance, voter turnout in a specific election is officially counted and reported. The questions for the experiment were chosen from a pool of questions that were used in previous studies, in particular in the three studies just cited above. The questions cover various fields of knowledge. The questions were chosen so that subjects were unlikely to know the exact answer. At the same time questions for which they did not have any knowledge at all were avoided. In order to avoid highly skewed responses the questions were such that the correct answers lay in an interval of 0% to 100%. The complete list of questions is

[7] Two participants had to be excluded from payment for the use of a mobile phone. Whether they intended to cheat in the task or used their phones for other purposes is not known. The experimental results do not rely on decisions of these two subjects. Their decisions are kept in the results reported in the paper.

reported in Table C.1. Participants could answer with any integer number between (and including) 0 and 100. The time to answer a question was not limited, the subjects were, however, given a reference time of 25 seconds per answer. The remaining time could be observed on the screens, but participants were informed beforehand that running out of time did not bear any consequences.

Phase	Identifier	Question	Correct Answer
A	A1	What was the voter turnout of the federal elections in Germany in 2005?	78
	A2	What is the share of water in a cucumber?	95
	A3	What share of the world-wide land area is used for agriculture?	18
	A4	What is the percentage of the world's population that lives in North- and Southamerica?	14
	A5	What is the percentage of the world's population between 15 and 64 years old?	65
	A6	What is the percentage of female professors in Germany?	18
	A7	What is the share of people with blood type B (BB or B0)?	11
	A8	What is the percentage of the world's roads (paved and unpaved) that are in India?	11
B	B1	What was the voter turnout of the federal elections in Germany in 2009?	71
	B2	What is the share of water in an onion?	89
	B3	What share of the working population is working in the agricultural sector?	40
	B4	What is the percentage of the world's population that lives in Africa?	15
	B5	What is the percentage of the world's population older than 15, that can read and write?	82
	B6	What is the percentage of female Nobel laureates in literature (until 2010)?	11
	B7	What is the share of people with blood type A (AA or A0)?	43
	B8	What is the percentage of the world's airports that are located in the United States?	30

Table C.1: Overview of all Questions

Confidence was measured on a nine point scale from 0 to 65+. Each value indicated a range of expected deviation of the individual estimate from the true value. The scale was explained in the instructions. For better understanding, a verbal interpretation was added. Table C.2 appeared in the instructions and gives a detailed description. As it can be seen in Table C.2, the distances between successive items are increasing. A simple nine point scale from 1 to 9 would have created the impression of equivalence of the distances. To avoid misinterpretation of the scale the values on the scale used in the experiment directly corresponded to the expected range of deviations. The same values were displayed on the screens. The confidence indication was not (directly) incentivized.

C.2 Phase I

The experiment consisted of two phases. The first set of instructions was handed out to the participants before the first phase. In phase I, each subject had to answer eight questions and indicate her confidence level. Each question was answered once. An English translation to the questions was provided on the screens. The order of the questions was randomized over the participants. The subjects were informed that there was going to

Table C.2: Summary of the Confidence Scale and its Interpretation

Scale	I assume that my estimation most likely (in nine of ten cases) does not deviate by more than	Concerning my estimation I am
0	0 percentage points from the true value	absolutely confident
1	1 percentage points from the true value	pronouncedly confident
2	2 percentage points from the true value	very confident
4	4 percentage points from the true value	rather confident
8	8 percentage points from the true value	partially confident
16	16 percentage points from the true value	rather unconfident
32	32 percentage points from the true value	very unconfident
64	64 percentage points from the true value	pronouncedly unconfident
65+	65 percentage points or more from the true value	absolutely unconfident

be a second phase with new instructions and that their choices in phase I might affect phase II. However, participants did not know the relation between phase I and phase II. In particular, they were given the instructions for phase II only after phase I was finished. Figure C.1 shows a screen shot of phase I.

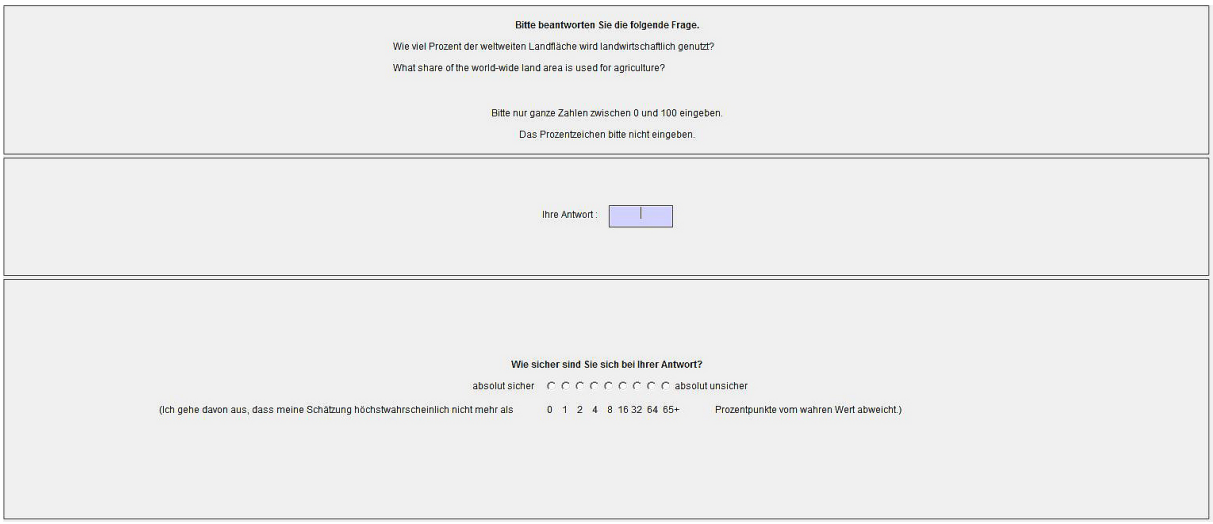


Figure C.1: Screen Shot of phase I

C.3 Phase II

Before phase II started, a new set of instructions was handed out and read aloud. The participants kept the first set of instructions. After participants were offered to ask

questions, which were again answered privately, phase II started. The participants were randomly matched into groups of four. Four is the minimum number of individuals required for a star network that is no simple line. The groups were fixed for the remainder of the experiment. The subjects were again asked to answer eight questions, but in phase II each question was answered six times ($t = 1, ..., 6$ estimation periods). In the first estimation period the subjects individually answered the questions and stated their confidence level. The first period was, therefore, analogous to phase I. In the second estimation period the subjects were informed about the other group members' guesses in period one according to their position in a star network. The subject at the central node could observe all answers and confidence information given by the members of her group. The pendants could observe the answer and confidence information given by their group's center. In addition, everyone was shown their own last answer and confidence statement. Participants then submitted new estimates and confidence levels. Period two was repeated four times with the information from the respective previous round. It was communicated in the instructions for both phases that the payment, although they were playing in groups, was based solely on the individual error. The order of the questions was randomized across groups. The subjects only had to wait for for their group members for each estimation period, but had to wait for all participants of the same session to enter a new question round. Figures C.2 and C.3 show screen shots of phase II.

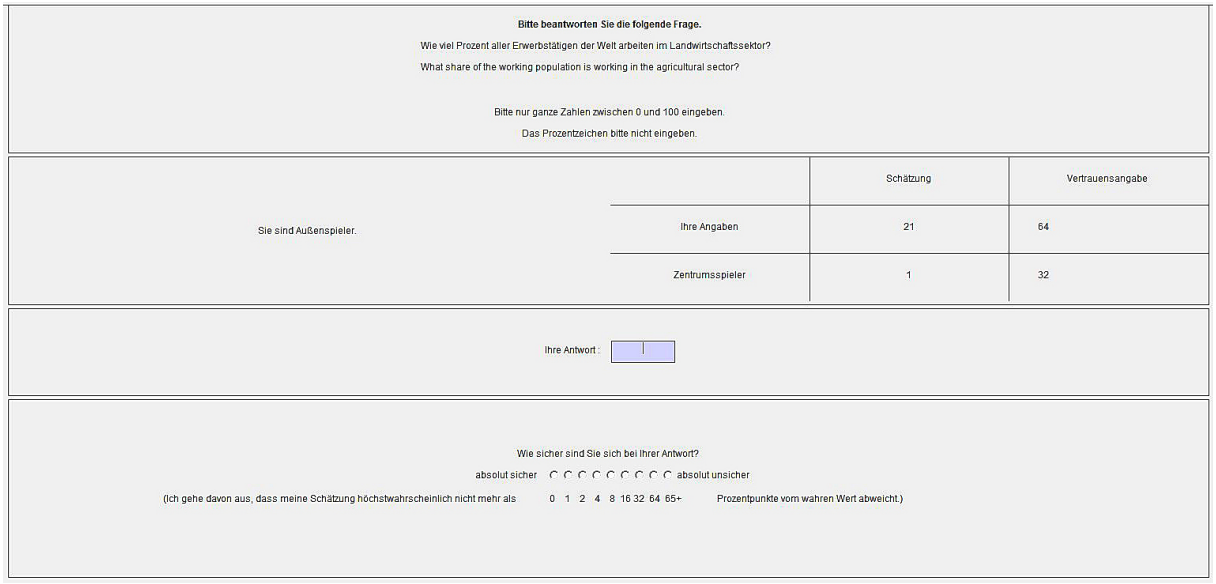


Figure C.2: Screen Shot of phase II from the Viewpoint of a Pendant

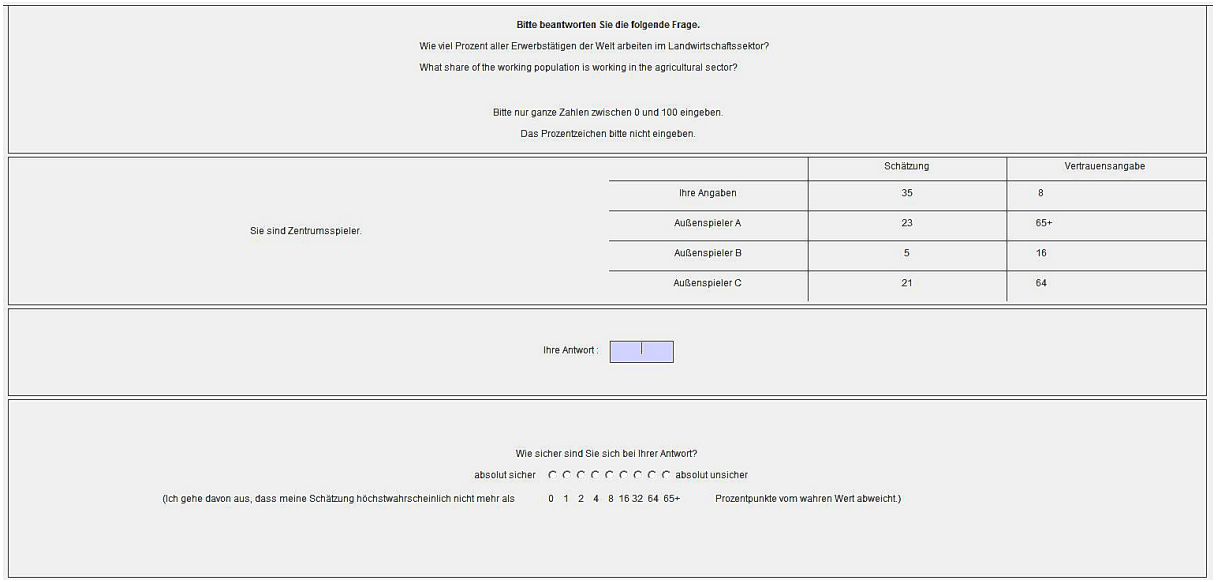


Figure C.3: Screen Shot of phase II from the Viewpoint of a Center

C.4 Treatments

The selection criterion for the individual at the center of the star network varied with the treatments. The center was fixed for one question round, that is for six estimation periods. The pendants were named *A,B* and *C*. The names changed with each question round. The positions and the selection criteria were communicated on screen at the beginning of each question round. The criterion changed once after the fourth question round. This piece of information was communicated to the subjects in the instructions for phase II.

The selection was either random or it was based on information from phase I. Each question in phase II had a partner question in phase I. The partner questions were two questions considered similar to each other and from the same field of interest. It could be expected that the individual error and confidence levels for the two partner questions were correlated. It was communicated that questions in phase II partially resembled questions from phase I, but were never identical.

Treatment T1 was a high accuracy treatment. This type of selection rule is based on the quality of the estimates of the partner questions in phase I. In each group, the agent who had made the smallest error in her estimation of the answer to the partner question was chosen to be in the center for this question round. Treatment T2 was a high confidence treatment. The selection in treatment T2 was based on the confidence level indicated in phase I for the partner question. The agent with the highest confidence within each group was selected. Treatment T0 was a control treatment. In this treatment the central agent was determined by a random pick. It was also made clear that in case of a tie in T1 or T2, a random choice was made between the group members who were

eligible for the center position.

It was necessary to separate the question determining the central node from the actual question of interest and, therefore, include two phases in the experiment for several reasons. If the selection had been based on the first estimation period, by communicating that the center had had the best initial guess on that question in treatment T1, the participants would have gained outside information about the true state of the world. With the introduction of the partner question it was possible to reduce the level of outside information to a minimum. Furthermore, the questions were asked in successive rounds, i.e., the second to sixth period of every question round immediately followed the initial period. Communication of the selection mechanism could have induced strategic considerations on the side of the participants. Particularly in treatment T2, a misrepresentation of the confidence level to get the desired position in the network could have been expected. In order to rule out intentional misrepresentation, we collected the data in phase I before its role in the determination of the network of phase II was communicated.

C.5 Permutation of Treatments

To avoid session effects, each treatment was run at each session. Each group played two different treatments in a fixed order. They started with either treatment T1 or T2 and then changed to treatment T0 after the fourth question round, or they started with treatment T0 and then changed to treatment T1 or T2. Table C.3 shows a detailed breakdown of the distribution of participants over the treatments.

Table C.3: Summary of Treatments

Treatment	Participants
T0 Random / T1 Accuracy	44
T0 Random / T2 Confidence	40
T1 Accuracy / T0 Random	44
T2 Confidence / T0 Random	48

C.6 Payment

The subject's payment was based on the individual error of the estimation. The error was calculated as the absolute difference between the estimation and the correct answer. The individual error was converted into game points as described in Table C.4.

Table C.4: Summary of Payments

Distance	Points
0 percentage points	16 game points
1 percentage point	8 game points
2 percentage points	4 game points
3 or 4 percentage points	2 game points
5,6,7 or 8 percentage points	1 game point
more than 9 percentage points	0 game points

One question was randomly selected for payment in phase I. In phase II, one estimation period was randomly chosen for each question round. The choice was identical for everyone. The monetary incentive in this form encouraged the subjects to find the true answers. Payoffs only depended on one's own decisions. Neither there is incentive to improve other's choices or to be better than others. The experimental design put subjects into a position in which they would try to get as close to the truth as possible by using their own knowledge and information from others (Lorenz et al., 2011). The game points were converted with an exchange rate of EUR 0.3 per point. The total payment was: EUR 5 show up fee + game points from phase I $\cdot$ 0.3 + game points from phase II $\cdot$ 0.3. The maximum payment possible was EUR 48.2.

The experiment was concluded by a short questionnaire. After all participants had finished answering the questionnaire, the correct solutions to all estimate questions were displayed on screen. The participants were then paid anonymously at two cash desks at the exits of the laboratory.

C.7 Pretest Sessions

Two pretest sessions were run on the 6th and 19th of August, 2013 with 15 and 16 participants, respectively, in order to calibrate the number of questions and the conversion rate. We determined the sample size prior to the experiment and the pretest sessions by a heuristic statistical power analysis.

D Instructions

The original instructions are written in English and in German. On the next pages we provide the original instructions, first for phase I of the experiment, then for phase II.

Herzlich Willkommen zum Experiment!

Sie nehmen nun an einem Experiment zur ökonomischen Entscheidungsfindung teil. Bitte beachten Sie, dass ab nun und während des gesamten Experiments **keine Kommunikation** gestattet ist. Wenn Sie eine Frage haben, strecken Sie bitte die Hand aus der Kabine, einer der Experimentatoren kommt dann zu Ihnen. Während des gesamten Experiments ist das Benutzen von Handys, Smartphones, Tablets oder Ähnlichem untersagt. Bitte beachten Sie, dass eine Zuwiderhandlung zum Ausschluss von dem Experiment und von sämtlichen Zahlungen führt.

Sämtliche Entscheidungen erfolgen anonym, d.h. keiner der anderen Teilnehmenden erfährt die Identität des Anderen. Auch die Auszahlung erfolgt anonym am Ende des Experiments. Das bedeutet, dass keiner der anderen Teilnehmenden erfährt, wie hoch Ihre Auszahlung ist.

Anleitung zum Experiment und allgemeine Informationen

Das Experiment besteht aus zwei Phasen. Sie erhalten zunächst die Instruktionen für die Phase I des Experiments. Die Instruktionen für die Phase II erhalten Sie nachdem alle Teilnehmenden die erste Phase abgeschlossen haben. Ihre Angaben in Phase I können in manchen Fällen Einfluss auf Phase II haben. Auf Phase II folgt ein kurzer Fragebogen.

Informationen zu Phase I des Experiments

In diesem Experiment geht es um das möglichst gute Einschätzen von bestimmten Größen. Je nach Qualität Ihrer Schätzungen erhalten Sie Punkte, die am Ende des Experiments zu Ihrer Auszahlung in Euro führen.

In der ersten Phase des Experiments werden Sie gebeten, acht Fragen zu beantworten. Gefragt ist jeweils nach einem Prozentwert und Sie können stets nur ganze Zahlen zwischen (und einschließlich) 0 und 100 als Schätzung angeben. Das Prozentzeichen soll dabei nicht eingegeben werden. Die wahren Werte beruhen auf offiziellen Statistiken und wurden, insofern dies nötig war, auf ganze Zahlen gerundet.

Sie werden außerdem gebeten, Ihr Vertrauen in Ihre Schätzung auf einer Skala anzugeben (Vertrauensangabe). Bitte entnehmen Sie die Bedeutung der Werte auf der Skala der folgenden Übersicht, die jedem Zahlenwert auch eine verbale Interpretation beifügt:

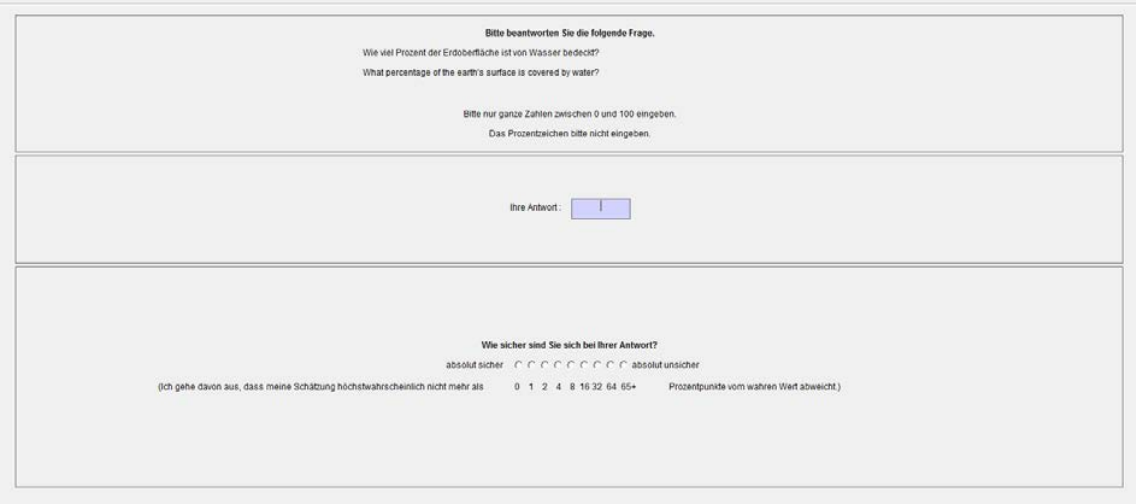
Skala	Ich gehe davon aus, dass meine Schätzung höchstwahrscheinlich* nicht mehr als	Ich bin mir bei meiner Schätzung:
0	0 Prozentpunkte vom wahren Wert abweicht.	absolut sicher
1	1 Prozentpunkt vom wahren Wert abweicht.	ausgesprochen sicher
2	2 Prozentpunkte vom wahren Wert abweicht.	sehr sicher
4	4 Prozentpunkte vom wahren Wert abweicht.	eher sicher
8	8 Prozentpunkte vom wahren Wert abweicht.	teilweise sicher
16	16 Prozentpunkte vom wahren Wert abweicht.	eher unsicher
32	32 Prozentpunkte vom wahren Wert abweicht.	sehr unsicher
64	64 Prozentpunkte vom wahren Wert abweicht.	ausgesprochen unsicher
65+	65 Prozentpunkte oder mehr vom wahren Wert abweicht.	absolut unsicher

(* in 9 von 10 Fällen)

Beispiel:

Nehmen wir an, Ihre Schätzung beträgt 50% und Sie sind sich bei dieser Schätzung „eher sicher“ (Skalenwert 4). Das bedeutet, dass Sie davon ausgehen, dass Ihre Schätzung von 50% höchstwahrscheinlich (in 9 von 10 Fällen) nicht mehr als 4 Prozentpunkte von dem wahren Wert abweicht, der wahre Wert also zwischen 46% und 54% liegt.

Grafik 1 zeigt beispielhaft, welche Bildschirmoberfläche Sie bei jeder Frage erwartet. In das Eingabefeld für die Schätzung, soll eine Zahl zwischen 0 und 100 eingegeben werden. Darunter erfolgt die Angabe des Vertrauens auf der angezeigten Skala. Bitte bestätigen Sie Ihre Eingaben durch Klick auf den Weiter-Button (nicht ersichtlich in *Grafik 1*).



Berechnung Ihres Einkommens aus Phase I

Grundlage für die Gewinnberechnung ist der Abstand Ihrer Schätzung zum richtigen Wert – je näher Sie am richtigen Wert liegen, desto mehr Geld erhalten Sie. Der Abstand wird berechnet als der absolute Betrag der Differenz zwischen Ihrer eigenen Schätzung und dem wahren Wert. Ihr Gewinn hängt ausschließlich von Ihrer eigenen Schätzung ab.

Punktevergabe:

- **16 Punkte** erhalten Sie, wenn Ihre Schätzung exakt den richtigen Wert trifft (**0 Prozentpunkte** Abstand).
- **8 Punkte** erhalten Sie, wenn Ihre Schätzung fast exakt den richtigen Wert trifft (**1 Prozentpunkt** Abstand).
- **4 Punkte** erhalten Sie für eine kleine Abweichung der Schätzung vom wahren Wert (**2 Prozentpunkte** Abstand).
- **2 Punkte** erhalten Sie für eine mittlere Abweichung der Schätzung vom wahren Wert (**3 oder 4 Prozentpunkte** Abstand).
- **1 Punkt** erhalten Sie für eine größere Abweichung der Schätzung vom wahren Wert (**5, 6, 7, oder 8 Prozentpunkte** Abstand).
- Weicht Ihre Schätzung stark vom richtigen Wert ab (Abstand von **9 Prozentpunkten und mehr**), erhalten Sie für diese Runde **keine Punkte**.

Für Ihr Einkommen in Phase I wird aus den 8 Fragen zufällig **eine** auszahlungsrelevante Frage ausgelost. Ihr Einkommen ergibt sich dann durch Ihre dort erzielte Punktzahl, wobei folgender Wechselkurs gilt: **1 Punkt entspricht 0,30 €.** Das maximal mögliche Einkommen in Phase I beträgt 4,80 €.

Beispiel (fortgesetzt):

Sie haben bei einer Frage 50% geschätzt. Nehmen wir an, dass der wahre Wert bei 48% liegt, dann beträgt Ihr Abstand 2. Wenn diese Frage als auszahlungsrelevant ausgelost wird, dann bekommen Sie 4 Punkte und damit 1,20 € ausgezahlt.

Gesamteinkommen

Ihr Gesamteinkommen aus dem Experiment setzt sich aus den garantierten 5 €, plus Ihrem Einkommen aus Phase I, plus Ihrem Einkommen aus Phase II zusammen und wird am Ende des Experiments ausgezahlt.

Viel Erfolg!

Welcome to today‘s experiment!

You are now participating in an experiment concerning economic decision making. Please note that from now on and during the time of the experiment **communication is not allowed**. If you have any questions, please indicate this by showing your hand outside of the individual cabin; one of the experimenters will come to assist you. The use of mobile phones, tablet PCs and similar devices is not allowed during the time of the experiment. Please note that a violation of this rule will lead to an expulsion of the experiment and will exclude you from any payment.

All decisions are made anonymously, i.e. none of the other participants will get to know the identity of a decision maker. Similarly, the payment is made anonymously such that none of the other participants will get to know how much you earn.

Instructions and general information

The experiment consists of two phases. You are now holding the instructions for phase I. You will receive the instructions for phase II after all participants have completed phase I. In some cases the choices in phase I might affect phase II. After phase II there will be a short questionnaire to answer.

Information about phase I of the experiment

This experiment is about estimating certain figures as accurately as possible. Your score depends on the quality of your estimations and will be transformed into a payment in Euros at the end of the experiment.

In the first phase of the experiment you are asked to answer eight questions. Each questions is about some percentage of a face value and you can type in integer numbers between (and including) 0 and 100 as an estimate. Thereby, the percent sign should not be typed in. The true values are based on some official statistical reports and were, if applicable, rounded to the next integer.

In addition, we ask you for your confidence in your estimate on a scale. The meaning of each value of the scale can be found in the following table, which adds a verbal interpretation to each quantity.

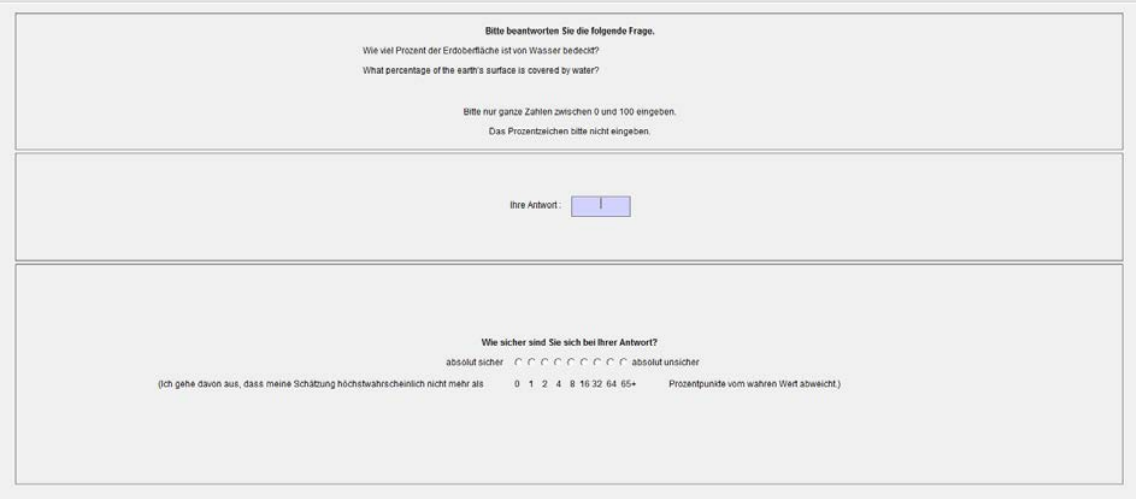
Scale	I assume that my estimation most likely* does not deviate by more than	Concerning my estimation I am
0	0 percentage points from the true value.	absolutely confident
1	1 percentage points from the true value.	pronouncedly confident
2	2 percentage points from the true value.	very confident
4	4 percentage points from the true value.	rather confident
8	8 percentage points from the true value.	partially confident
16	16 percentage points from the true value.	rather unconfident
32	32 percentage points from the true value.	very unconfident
64	64 percentage points from the true value.	pronouncedly unconfident
65+	65 percentage points or more from the true value.	absolutely unconfident

(* in 9 out of 10 cases)

Example:

Suppose your estimation is 50% and concerning this estimation you feel „rather confident“ (value 4 on scale). This means that you assume that your estimation most likely (in 9 out of 10 cases) does not deviate by more than 4 percentage points from the true value, i.e. that the true value lies within 46% and 54%.

Figure 1 gives an example for the screen which you will see for each question. The first input is your estimation, which must be a number between 0 and 100. The second input is your confidence level. Please confirm your choices by clicking on the „Weiter“ button (which is not illustrated in *Figure 1*).



Calculation of your income from phase I

Profits are based on the distance of your estimation to the true value – the closer you are to the true value, the more money you earn. The distance is computed as the absolute value of the difference between your estimation and the true value. Your profit solely depends on your own estimation.

Score:

- You receive **16 points** if your estimation hits exactly the true value (distance of **0 percentage points**).
- You receive **8 points** if your estimation hits almost exactly the true value (distance of **1 percentage point**).
- You receive **4 points** for a small distance of your estimation to the true value (distance of **2 percentage points**).
- You receive **2 points** for a medium distance of your estimation to the true value (distance of **3 or 4 percentage points**).
- You receive **1 point** for a larger distance of your estimation to the true value (distance of **5, 6, 7, or 8 percentage points**).
- If your estimation strongly deviates from the true value (distance of **9 percentage points or more**), then you receive **no points** in this round.

To generate your income of phase I, **one** of the 8 questions will be randomly drawn to be payoff-relevant. Your income in phase I is then derived from your score in this question, whereas the following exchange rate applies: **1 point corresponds to 0,30 €**. The maximal possible income in phase I is 4,80 €.

Example (continued):

You have estimated 50% in some question. Let us suppose that the true value is 48%. Then your distance is 2. If this question is drawn to be payoff-relevant, then you receive 4 points such that you will earn 1,20 €.

Total income

Your total income from the experiment consists of the guaranteed 5 €, plus your income from phase I, plus your income from phase II, and will be paid by the end of the experiment.

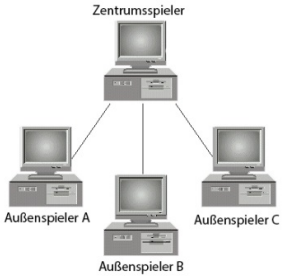
Good luck!

Informationen zu Phase II des Experiments:

In Phase II werden Sie erneut gebeten, 8 unterschiedliche Fragen zu beantworten und Ihre jeweilige Vertrauensangabe zu machen. Bitte beachten Sie, dass die Fragen in Phase II den Fragen aus Phase I des Experiments teilweise ähnlich sind, sie sind jedoch in keinem Fall identisch! Nach Ihrer ersten Schätzung zu einer Frage werden Sie noch fünf weitere Male gebeten, eine Schätzung für die gleiche Frage abzugeben. Ab der ersten Wiederholung bekommen Sie je nach Spielmodus Informationen über die Angaben anderer Spieler.

Zu Beginn der Phase II werden Sie entweder in eine Vierergruppe eingeteilt oder Sie spielen diese Phase einzeln. Sowohl die Zuordnung als auch die Zusammensetzung der Gruppen erfolgen zufällig und ändern sich während des Experiments nicht mehr.

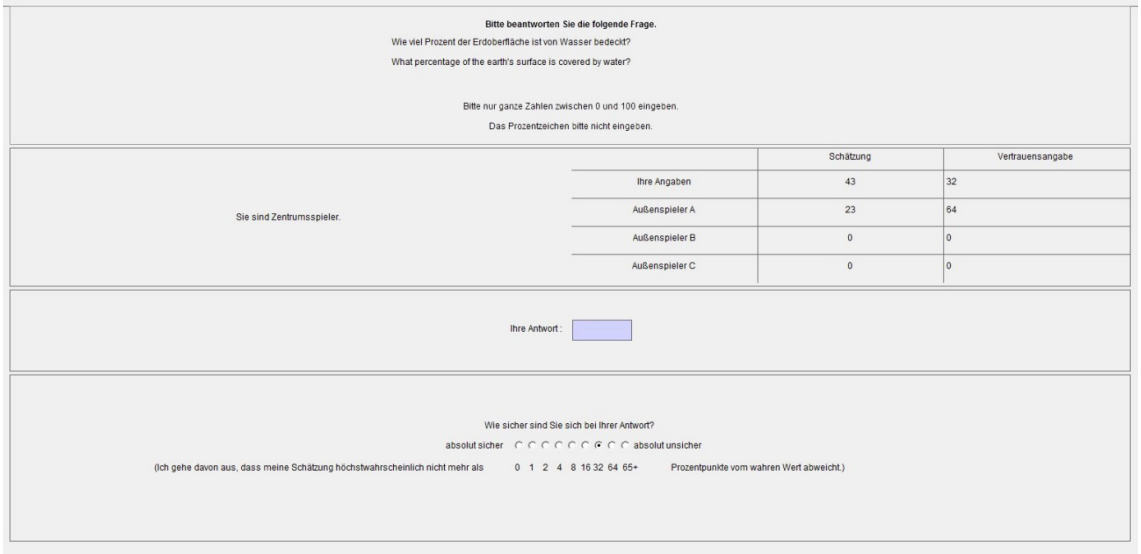
Gruppenmodus



In jeder Vierergruppe wird ein Spieler für die Dauer einer Frage für die Rolle des **Zentrumsspielers** ausgewählt, während die drei anderen die **Außenspieler** sind *(Grafik 2)*. Der Auswahlmechanismus wird jeweils bekannt gegeben und wechselt einmal nach der vierten Frage. Die Auswahl basiert entweder auf Angaben aus Phase I oder erfolgt zufällig.

Sie werden nun sechs Mal geben, eine Schätzung für die gleiche Frage abzugeben. In der ersten Schätzrunde stehen noch keine Informationen zur Verfügung. Von der zweiten bis zur sechsten Schätzrunde sieht der Zentrumsspieler die vorangegangenen Schätzungen und Vertrauensangaben der Außenspieler, während die Außenspieler die Schätzung und Vertrauensangabe des Zentrumsspielers sehen, nicht jedoch die Eingaben der jeweils anderen Außenspieler.

Grafiken 3 und 4 zeigen beispielhaft, wie die Bildschirmoberflächen für einen Außenspieler und einen Zentrumsspieler aussehen können.



Bitte beachten Sie, dass die Bezeichnungen A, B und C für jede Frage, also für sechs Antworten, bestehen bleiben und dann neu vergeben werden.

Im **Einzelmodus** werden ebenfalls 6 Schätzungen zu jeder Frage abgegeben. Die dabei bereitstehenden Informationen werden für jede Frage auf dem Bildschirm erläutert. Die Art der Information wechselt einmal nach der vierten Frage.

Berechnung Ihres Einkommens aus Phase II

Für die Berechnung des Einkommens aus Phase II wird für **jede** der 8 Fragen genau **eine** auszahlungsrelevante Runde zufällig durch den Computer bestimmt. Genau wie in Phase I ist der Abstand Ihrer Schätzung zum wahren Wert Grundlage für die Gewinnberechnung. Je näher Sie in der zufällig ausgewählten Runde am richtigen Wert liegen, desto mehr Geld erhalten Sie (siehe Punktevergabe in den Instruktionen zu Phase I). Bitte beachten Sie, dass, auch wenn Sie in einer Gruppe spielen, nur **Ihre eigene Schätzung** Einfluss auf Ihren Gewinn hat.

Ihr Einkommen aus Phase II ergibt sich aus der Summe der Punkte, die Sie für jede Frage in der jeweils auszahlungsrelevanten Runde gesammelt haben, wobei nach wie vor der Wechselkurs von **1 Punkt entspricht 0,30 €** gilt. Das maximal mögliche Einkommen in Phase II beträgt 38,40 €.

Gesamteinkommen

Ihr Gesamteinkommen aus dem Experiment setzt sich aus den garantierten 5 €, plus Ihrem Einkommen aus Phase I, plus Ihrem Einkommen aus Phase II zusammen.

Ihre Auszahlung sowie die tatsächlichen Werte erfahren Sie am Ende des Experiments.

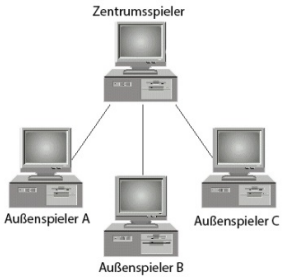
Viel Erfolg!

Information concerning phase II of the experiment

In phase II you are asked again to answer 8 questions and to provide your confidence levels. Please note that the questions of phase II partially resemble the questions of phase I, but they are never identical! After your first estimation concerning one question you will be asked 5 further times to provide an estimation for the same question. After the first repetition – depending on the mode of play – you will receive information about other players' decisions.

At the beginning of phase II you are either assigned into a group of four players or you will be a single player. Both the assignment and the composition of the groups are generated randomly and will not change for the time of the experiment.

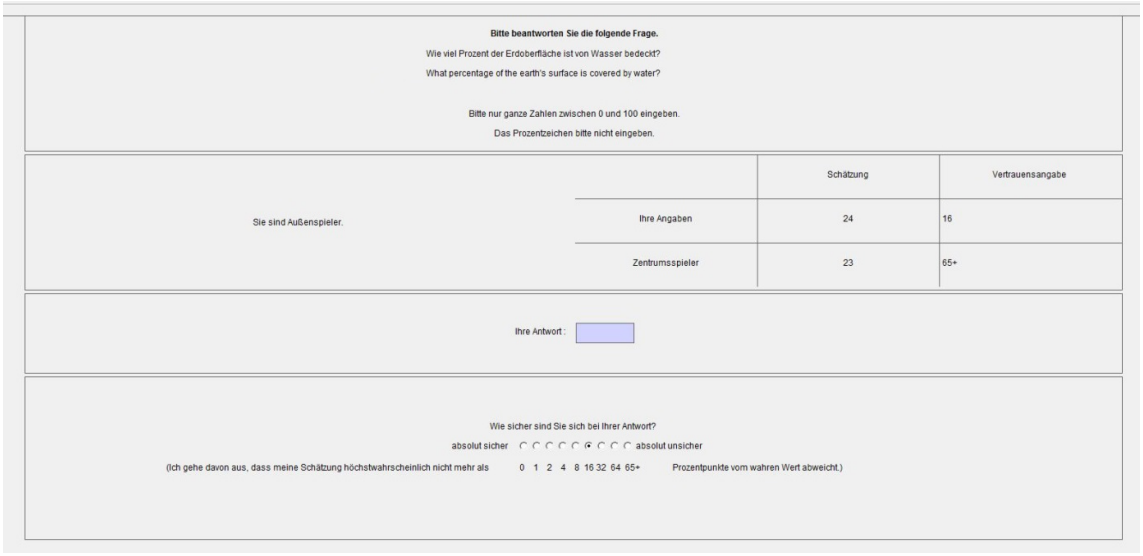
Group mode

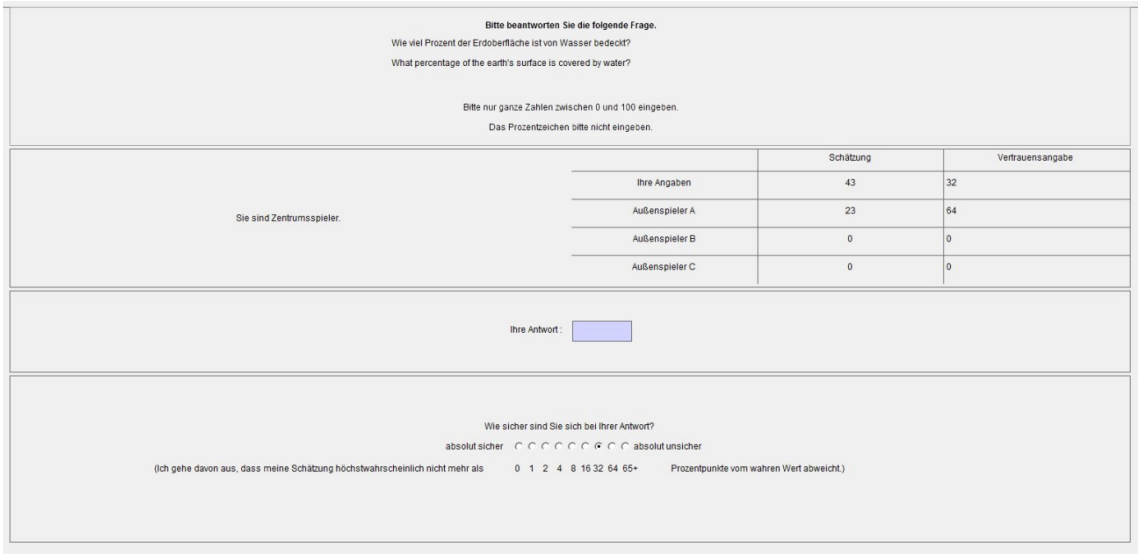


In each group of four, one player is selected to be the **central player** for the time of one question, while the other three are **peripheral players** *(Figure 2)*. The selection mechanism is announced each time and will once change after four questions. The selection is either based on inputs from phase I or is made randomly.

You will then be asked 6 times to provide an estimation for the same question. In the first round of estimation no information is provided. From the second round up to the 6th round the central player can see the previous estimations and confidence choices of the peripheral players, while the peripheral players can see the previous estimation and confidence choice by the central player, but not those of the other peripheral players.

Figure 3 and 4 illustrate how computer screens of a peripheral player and of a central player might look like.





Please note that the labels A, B, and C are fixed for each question, i.e. for six answers, and then they are newly assigned.

In the **single-player mode** you also have to provide 6 estimations for each question. The available pieces of information will be specified for each question on the computer screen. The type of information changes once after the fourth question.

Calculation of your income in phase II

To compute your income in phase II, for **each** of the 8 questions **one** payoff-relevant round will be randomly selected by the computer. Exactly as in phase I, your profit is based on the distance of your estimation to the true value. The closer you are to the true value, the more money you earn (see definition of score in instructions of phase I). Please note that, even if you play in group mode, **solely your own estimation** affects your profit.

Your income from phase II is derived from the sum of points you have collected for each question in the corresponding payoff-relevant round, whereas the exchange rate is still **1 point corresponds to 0,30 €**. The maximal possible income in phase II is 38,40 €.

Total income

Your total income from the experiment consists of the guaranteed 5 €, plus your income from Phase I, plus your income from phase II.

Your payoff as well as the correct answers will be provided by the end of the experiment.

Good luck!

NOTE DI LAVORO DELLA FONDAZIONE ENI ENRICO MATTEI
Fondazione Eni Enrico Mattei Working Paper Series

Our Working Papers are available on the Internet at the following addresses:
http://www.feem.it/getpage.aspx?id=73&sez=Publications&padre=20&tab=1

NOTE DI LAVORO PUBLISHED IN 2018

1. 2018, CSI Series, Claudio Morana, Giacomo Sbrana, Some Financial Implications of Global Warming: an Empirical Assessment

2. 2018, ET Series, Berno Büchel, Stefan Klößner, Martin Lochmüller, Heiko Rauhut, The Strength of Weak Leaders - An Experiment on Social Influence and Social Learning in Teams

Fondazione Eni Enrico Mattei
Corso Magenta 63, Milano – Italia

Tel. +39 02.520.36934
Fax. +39.02.520.36946

E-mail: letter@feem.it
www.feem.it

FONDAZIONE ENI
ENRICO MATTEI