Learning Through Imitation: An Experiment*

Marina Agranov† Gabriel Lopez-Moctezuma‡
Philipp Strack§ Omer Tamuz¶

April 11, 2022

Abstract

We compare how well agents aggregate information in two repeated social learning environments. In the first setting agents have access to a public data set. In the second they have access to the same data, and also to the past actions of others. Despite the fact that actions contain no additional payoff relevant information, and despite potential herd behavior, free riding and information overload issues, observing and imitating the actions of others leads agents to take the optimal action more often in the second setting. We also investigate the effect of group size, as well as a setting in which agents observe private data and others' actions.

*This work was presented at Princeton University, MiddExLab virtual seminar, Chicago Harris, UCSB, NYU, University of Michigan, Osaka University, Texas A&M, Stanford University, CREST, University of Hamburg, University of Maryland, Caltech, Essex University, NYU Abu Dhabi, University of Southamptons, Kansas University, Shanghai Jiao Tong University, and GAMES 2020. We thank the participants of these seminars and conferences and Carlo Cusumano for their helpful comments.

†Caltech. Email: magranov@hss.caltech.edu. Experiments were funded by the grant received from the Ronald and Maxine Linde Institute of Economic and Management Sciences at Caltech.

‡Caltech. Email: glmoctezuma@caltech.edu.

§Yale University. Email: philipp.strack@yale.edu.

¶Caltech. Email: tamuz@caltech.edu. Omer Tamuz was supported by a grant from the Simons Foundation (#419427), a Sloan fellowship, a BSF award (#2018397) and a National Science Foundation CAREER award (DMS-1944153).

1 Introduction

A large literature in economics and finance has shown that imitation in social learning leads to inefficiencies: agents who imitate do not reveal their private information, giving rise to information cascades and herd behavior. This point has been made both theoretically and experimentally in a variety of settings, including those with rational agents (Banerjee, 1992; Bikhchandani et al., 1992; Anderson and Holt, 1997; Çelen and Kariv, 2004; Harel et al., 2021) and those that consider behavioral biases and heuristics (Enke and Zimmermann, 2019; Eyster et al., 2018).

We study a social learning setting in which all information is *public*, and ask whether agents perform better or worse when given the opportunity to imitate the actions of others. Although these actions do not contain additional payoff relevant information beyond the public raw data, and despite their potential to create herd behavior, free riding, and information overload, we find that agents perform better when allowed to observe others' actions.

In our experiment, participants play the following game: An urn is chosen to be either majority green or majority red, each with probability one half. In the former case the urn contains six green balls and four red balls, whereas in the latter case the numbers are reversed. A signal is a random draw (with replacement) of a ball from the urn. A game consists of twenty periods. In each period, each member of a group of eight participants observes some information and then has to guess the majority color of the urn. Participants are rewarded for guessing correctly.

In each period, each participant observes a signal. Furthermore, depending on the treatment, participants observe some additional information. In the SIGNALS treatment, participants observe others' signals. In the ALL treatment, they likewise observe all the others' signals, and also all the others' past actions. Thus, in both treatments all signals are public, but actions are public only in the latter.

The comparison between the SIGNALS and ALL treatments sheds light on the usefulness of observing other people's decisions, in addition to the raw public data. Rational agents are expected to perform identically in these two treatments, since both offer agents exactly the same information about the state; in both cases all agents can see all the signals available to society. Some behavioral heuristics suggest that agents should do better when observing only signals, as observing the actions of others can—through information overload and correlation neglect—lead to wrong conclusions, potentially through herd behavior and groupthink. Information overload has been known to interfere with information processing and may lead to sub optimal decisions (Caplin et al., 2011; Scheibehenne and Greifeneder, 2010). Since the ALL treatment features twice as many pieces of information as the SIGNALS treatment, these phenomena suggest that performance in the ALL treatment should be worse. Furthermore, the literature has documented a tendency of people to neglect correlation among dependent pieces of information (Enke and Zimmermann, 2019). In the SIGNALS treatment, correlation neglect plays no role because the signals of other group members are independent conditional on the state of nature. However, in the ALL treatment, correlation neglect might be detrimental since others' actions are not independent and regarding them as independent will hinder learning. Finally, pure imitation by all agents is clearly inefficient, as it leaves no room for the signals to inform the actions.

Nevertheless, we find that agents perform better in the ALL treatment than in the SIGNALS treatment: they are better off when they are privy to others' actions, in addition to the information that those actions are based on. A possible explanation for this finding is that observing the actions of others allows subjects to benefit from the aggregation of information done by other subjects. We observe that subjects rely on the actions of others more heavily when signals are "weak", by which we mean relatively uninformative (i.e., roughly the same number of red and green balls have been drawn) and it is thus harder to determine which color was drawn more.

We observe substantial heterogeneity in how well subjects aggregate raw data. This allows worse-performing subjects to benefit from the aggregation labor done by the better performing ones. In particular, subjects with higher IQ perform better in both treatments, in line with the intuition that aggregating information is a cognitively demanding task.[1] Nevertheless, both high IQ and low IQ subjects rely on others to a similar extent, and do so almost exclusively when signals are weak. Finally, both groups benefit from this practice: within each group, subjects who imitate others perform better than those who do not.

Having documented the benefits of imitation in repeated social learning setting with *public* information, we complement our findings by studying a similar setting with *private* information. We focus on the effects of information structures and group size

[1] We note here that while it might be hard for a subject to understand how they should aggregate the information available to them we have deliberately chosen a setting where the aggregation is computationally easy as it only requires determining the color of the majority of signals.

and conduct two additional treatments. In the NO INFO treatment each participant observes her own signals only, and does not see others' signals or actions. In the ACTIONS treatment, subjects again observe their own signals and not the signals of others, but do observe the past actions of others.

The comparison between the NO INFO and ACTIONS treatments reveals to what extent people are capable of extracting the information contained in their peers' private signals by observing their actions, in a repeated setting. Theory predicts that such inference is difficult and has limited benefits (Harel et al., 2021), which leads to the natural question of how well people perform in practice in such settings.

We find that participants perform significantly better in the ACTIONS treatment as compared with the NO INFO treatment, despite the complexity of the information extraction problem. In addition, participants rely on the actions of others more heavily when their private signals are relatively uninformative. This is reminiscent of the pattern observed with public signals.

Finally, we compare how changing the group size affect results, by repeating these experiments with groups of only four agents. While in the ALL treatment larger groups perform better, in the ACTIONS treatment performance is identical across group sizes. This result qualitatively matches a theoretical result due to Harel et al. (2021), who show that myopic Bayesian agents, in the long-run, do not perform significantly better in larger groups.

We proceed as follows. Related literature is surveyed in Section 2. Section 3 contains the detailed description of our experiment and Section 4 outlines theoretical predictions for each of the treatments. Section 5 reports the experimental results: Section 5.1 focuses on setting with public information (ALL versus SIGNALS treatments) and Section 5.2 considers settings with private information (the ACTIONS treatment) and documents group size effects. Finally, we offer some discussion of implications of our experimental findings in Section 6.

2 Related Literature

Our paper contributes to a large literature on social learning. Most of the theoretical literature focuses on sequential settings, in which exogenously ordered agents act once after observing some statistics about their predecessors' actions and their own private signal (see for example Banerjee, 1992; Bikhchandani et al., 1992; Smith and Sørensen, 2000, or a recent survey by Bikhchandani et al., 2021).

The experimental literature on sequential social learning is vast, and a complete survey is beyond the scope of this paper. The first experimental paper that documents informational cascades is Anderson and Holt (1997). Since then, a fruitful literature in experimental economics has studied the determinants and the limitations of herding behavior in sequential settings. Hung and Plott (2001) report the robustness of informational cascades under different rewards systems. Çelen and Kariv (2004) elicit subjects' beliefs in order to identify the source of imitation behavior. Kubler and Weiszacker (2004) allow subjects to purchase a private signal for a small fee. Angrisani et al. (2005) embed this game in a financial market setting. Goeree, Palfrey, Rogers, and McKelvey (2007) look at longer horizons games. Ziegelmeyer et al. (2010) study the fragility of information cascades patterns using groups of differently informed agents. Weizsacker (2010) conducts a meta-study of sequential social learning experiments and finds that subjects follow their own private information more frequently than is empirically optimal.[2] Duffy, Hopkins, Kornienko, and Ma (2019) allow subjects to choose between private and public information prior to guessing the state, while Duffy, Hopkins, and Kornienko (2020) in addition vary the persistence of the state across periods.

In comparison to this work, the repeated social learning game we study is quite different from the sequential move games surveyed above as, in our case, the same group of agents interacts repeatedly. This setting allows us to study how the information and the actions of others impact the beliefs and actions in an arguably more realistic setting. Accordingly, our paper is related to a new and exciting experimental literature that compares social learning outcomes under different network structures (Mueller-Frank and Neri, 2015; Grimm and Mengel, 2020; Chandrasekhar, Larreguy, and Xandri, 2020; Dasaratha and He, 2021; Choi, Gale, and Kariv, 2012). These papers vary the connections between agents in a group, i.e., who can observe whose actions, and ask how well subjects aggregate dispersed private information through repeated observations of neighbors' actions. Contrary to this literature, we fix the network

[2] Ziegelmeyer, March, and Krugel (2013) enlarge the data and use a modified methodology compared with Weizsacker (2010). They find that subjects tend to over-weigh their private signals but do so to a lesser degree than previously found. De Filippis et al. (2022) identify that such over-weighting occurs only when private signals are in conflict with social information. See also Angrisani et al. (2021) who study social learning in a continuous action space experiment and disentangle different theories that deliver over-weighting of private information.

structure to be complete, and vary the type of information our agents observe.[3]

Two recent laboratory games deserve special attention. The first one is Eyster, Rabin, and Weizsacker (2018) which compares social learning outcomes in a standard sequential move game and a cleverly constructed four-at-a-time move game. The sequential move game features participants who move one at a time after observing their predecessors' actions. The four-at-a-time game has four participants move in each period after observing predecessors from previous periods. Theoretically, Bayesian inference prescribes subjects to anti-imitate their predecessors in the latter but not the former game. The experimental data shows that subjects rarely anti-imitate in both games, which is at odds with the rational model. The four-at-a-time move game is interesting as it is one of the first environments in which, due to the incorrect processing of social information by participants, social learning is harmful on average, i.e., participants would do much better if they just ignored their predecessors actions and focused on their own private information instead.

The second paper is Evdokimov and Garfagnini (2020) who focus on groups with two players and vary whether both players, only one, or no one observes others' actions in a repeated social learning game. Similarly to our game, each agent gets a conditionally independent private signal in each period and reports her best guess about the state. However, contrary to our main results, the authors find no significant differences in the average quality of guesses across informational treatments. This difference between our papers indicates that group size plays an important role when aggregating social information.

From a theory perspective, our NO INFO, SIGNALS and ALL treatments are trivial to analyze, as they are equivalent to a single agent problem for a Bayesian decision maker. In contrast, the ACTIONS treatment is difficult to analyze within a rational framework, because higher order beliefs play a significant role. We elaborate on this in Section 4. Vives (1993) studies the speed of learning in a similar setting, but with a continuum of agents and continuous signals. Harel et al. (2021) study a setting that is identical to ours, under the assumption that agents are Bayesian and myopic. They focus on the long-term rate of learning, and show that even if many more agents are added to the group, the speed of learning remains bounded. Huang et al. (2021) extend this result to a network setting and to forward-looking agents.

[3]Moreover, in our game, participants get private signals in each game round, while in the papers above private signals are distributed only once at the very beginning of the game.

3 Experimental Design

Subjects are presented with the following scenario. There are two possible urns: a "red" urn and a "green" urn. The red urn contains six red and four green balls, and the green urn contains six green and four red balls. The color of the urn represents the state of nature (the superior policy, the best candidate for a job, the tastier croissant, etc.), which subjects are trying to learn by interacting repeatedly throughout the game. The structure and a sample of instructions are presented in sections A and B of the Online Appendix, respectively.

In all experiments, subjects play 10 games. At the start of each game, subjects are randomized into a group of eight subjects and one of the urns is chosen by a toss of a fair coin. This urn then remains unchanged for the duration of the game, and is used by all the participants of the game. Each game consists of 20 rounds. In each round, a subject guesses the color of the urn. We refer to these guesses as her actions. She then receives an independent draw (with replacement) from the chosen urn. The color of the drawn ball matches the color of the urn with probability 60%, which is therefore the precision of one's private signal; we purposefully chose an environment with low precision signals because in this environment the signals and actions of others are particularly valuable.

Depending on the treatment, subjects may observe additional information at the end of each round, which may help them choose their actions for the follow-up rounds. At the end of a session, a random round of a random game is selected uniformly. Subjects are rewarded for guessing correctly in that round: the correct action earns $20, while the incorrect one earns $5.

Our main variation between treatments is the information available to subjects at the time they choose their actions (in addition to their own signals). We consider four information structures:

(1) The NO INFO treatment, in which subjects observe only their private signals in every round of a game.

(2) The ACTIONS treatment, in which in addition to observing their own private signals, subjects also get to see the actions chosen by their group members in all previous rounds of the game.

(3) The SIGNALS treatment, in which subjects observe both their own private signals

and the private signals of their group members in all previous rounds of the game.

(4) The ALL treatment, in which subjects observe their own private signals, their group members' private signals and the actions chosen by their group members in all previous rounds of the game.

We conduct two additional treatments with the same information structure as the ACTIONS and the ALL treatments, but with smaller groups of four members each. We call these treatments ACTIONS4 and ALL4. These treatments allow us to explore how fast small and large groups learn depending on the information available to their members. Note that when subjects take their first action in round 1, they have no additional information about the state except for the prior. This sequencing of events within a round allows for a clean comparison between information treatments as we describe below, where all the information is delivered at the end of the round before the next round actions are taken.

Throughout the game, subjects have access to a table that keeps track of all signals and actions they observed in the past rounds of the game. This table presents information in an intuitive and visual way. This feature of the design ensures that our results are not affected by the memory subjects may have about the events that transpired during the game (see section B in the Online Appendix for screenshots).

At the end of the experiment, we elicit subjects' strategies with a series of open-ended questions as well as their beliefs about the fraction of correct actions of the other participants in the last round of the game in various treatments. Subjects are paid for the accuracy of their prediction in one randomly selected belief question. Subjects also report their gender and major, and complete a series of control tasks including risk attitudes elicitation using two investment tasks (Gneezy and Potters, 1997), IQ questions (ICAR, Condon and Revelle, 2014), and overconfidence.[4] In the analysis of experimental data, we classify subjects into *low-IQ* and *high-IQ* based on their answers to six IQ questions. We use their self-described strategies and beliefs to measure how useful low-IQ and high-IQ subjects think observing different types of information is, and whether it relates to their behavior in the experiment.

Table 1 summarizes the experimental design and number of participants. Experi-

[4]See sections B.2, B.3 and B.4 in the Online Appendix for details on the strategies, beliefs and other controls, respectively.

Table 1: Experimental Design

	Group size	Information					
		signals		actions			
Treatment		own	others	own	others	# of sessions	# of subjects
NO INFO	1	yes	no	yes	no	4 sessions	80 subjects
ACTIONS	8	yes	no	yes	yes	8 sessions	136 subjects
SIGNALS	8	yes	yes	yes	no	3 sessions	82 subjects
ALL	8	yes	yes	yes	yes	8 sessions	152 subjects
ACTIONS4	4	yes	no	yes	yes	4 sessions	76 subjects
ALL4	4	yes	yes	yes	yes	4 sessions	80 subjects
Total						31 sessions	606 subjects

mental sessions were conducted at two locations: University of California in San Diego (UCSD) and the Ohio State University (OSU).[5] The experiment was programmed and conducted with the oTree software (Chen et al., 2016) and was pre-registered in the AEA RCT registry (AEARCTR-0003315). Overall, 606 subjects participated in 31 sessions, and no subject participated in more than one session. The experiment lasted about 90 minutes. Subjects earned on average \$25.7, including a \$7 participation fee.

4 Theoretical Preliminaries

We model each game played by subjects in the experiment as follows. Denote by $\omega \in \{R, G\}$ the state of nature, which is chosen uniformly at random and stays fixed throughout each game. Denote by N the set of players. At the beginning of each round $t \in \{1, \ldots, 20\}$, each player i chooses an action $a_{i,t} \in \{R, G\}$. She then observes a random signal $s_{i,t} \in \{R, G\}$ such that conditioned on the state ω, the probability that the signal equals the state (i.e., $s_{i,t} = \omega$) is 60%. Conditional on the state, the signals are independent, both between time periods and players. Depending on the treatment, the players observe additional information; we denote by $I_{i,t}$ all that player i has observed before choosing her action $a_{i,t}$ at the beginning of round t. We use the terms action and guess interchangeably throughout the text.

[5]Because of closure of physical labs due to COVID-19, our data collection process at UCSD was interrupted and we had to finish collecting the data online at OSU using the same subject pool of students who normally participate in laboratory experiments. NO INFO, ACTIONS, and 4 sessions of ALL treatments were conducted in the experimental laboratory at UCSD, while the remaining 4 session of ALL treatment and SIGNALS treatment were conducted online at OSU. In section C of the Online appendix, we compare data collected at the physical lab at UCSD and in the online lab at OSU for the ALL treatment. We find that aggregate outcomes and individual level behavior of subjects in these two types of sessions are very similar to each other and statistically indistinguishable.

Each player plays some number of games. At the conclusion of player i's session one of these games is chosen uniformly at random, and in that game a round t is chosen uniformly at random. Player i receives a payoff equal to \$20 if she chose the correct action $a_{i,t} = \omega$, and to \$5 otherwise. We assume that players are Bayesian and that they have a strictly increasing utility for money. Under this assumption, behavior depends on the information players observe, but *not* on their preferences over risks.[6]

The NO INFO **and** SIGNALS **Treatments.** In these treatments the information $I_{i,t}$ available to player i in round t consists of a collection of signals that she observed up until this round. Specifically, in the NO INFO treatment this collection consists of her own signals only, $I_{i,t} = (s_{i,\tau})_{\tau<t}$, while in the SIGNALS treatment this collection contains signals of all group members, $I_{i,t} = (s_{j,\tau})_{j\in N,\tau<t}$. A standard calculation shows that the action $a_{i,t}$ is optimal if it is equal to any color that appears in $I_{i,t}$ at least as often as the other. In other words, the optimal action $a_{i,t}$ is equal to the majority color observed in the past signals, and can be both in case of a tie. Note the distinction between the correct action, i.e., the one that matches the state, and the optimal action, which is the one that matches the majority of the observed signals.

Given that subjects observe eight times more signals in the SIGNALS treatment than in the NO INFO treatment, theory predicts that subjects should be better at guessing the state in the former case in every round of the game (except the first, where they have no information). In particular, given our parameters, in the second round of the game, subjects should guess the state correctly 60% of the time in the NO INFO treatment and 71% of the time in the SIGNALS treatment. In the last round of the game, subjects are expected to guess the state 81% of the time in the NO INFO treatment and 99% of the time in the SIGNALS treatment. In Table 2 we provide the predicted probability of choosing correctly in each round.

The ALL **Treatment.** In the ALL treatment, subject i in round t observes all the signals and actions of her group members up until round t, so that $I_{i,t} = (s_{j,\tau}, a_{j,\tau})_{j\in N,\tau<t}$. Since the past actions $(a_{j,\tau})$ can contain no further information than the signals $(s_{j,\tau})$, in this case too optimal behaviour implies that $a_{i,t}$ is equal to any color that appears

[6] As there are only two outcomes, the possible payoff distributions are ranked in first order stochastic dominance. Therefore, the player will choose the same strategy under any preference that is monotone in first order stochastic dominance (e.g., cumulative prospect theory, etc.), as long as she forms beliefs using Bayes rule.

Table 2: Probabilities of Correct Actions

t	NO INFO	ACTIONS	SIGNALS/ALL	ACTIONS4	ALL4
1	0.50	0.50	0.50	0.50	0.50
2	0.60	0.60	0.71	0.60	0.65
3	0.60	0.73	0.79	0.68	0.71
4	0.65	0.77	0.84	0.71	0.75
5	0.65	0.78	0.87	0.73	0.79
6	0.68	0.80	0.90	0.75	0.81
8	0.71	0.83	0.93	0.77	0.86
10	0.73	0.84	0.96	0.79	0.89
12	0.75	0.86	0.97	0.81	0.91
14	0.77	0.87	0.98	0.83	0.93
16	0.79	0.88	0.99	0.84	0.94
18	0.80	0.89	>0.99	0.85	0.95
20	0.81	0.90	>0.99	0.86	0.96

Notes: Predicted probability of correct actions for parameters used in the experiment. In the ACTIONS treatments we assume common knowledge of rationality and that players act myopically.

in $(s_{j,\tau})_{j\in N,\tau<t}$ at least as often as the other. In other words, for Bayesian subjects we expect identical results in the ALL and SIGNALS treatments, since subjects should ignore others' actions and base their decisions on the observed signals.

The ACTIONS Treatment. The situation is more complicated in the ACTIONS treatment. Here, the information available to subject i in round t consists of her private signals and of the actions of the other subjects in all previous rounds, i.e., $I_{i,t} = (s_{i,\tau}, a_{j,\tau})_{j\in N,\tau<t}$. To the best of our knowledge, this case is analytically intractable. Nevertheless, a simulation of equilibrium behaviour is possible under the assumption that players act myopically, i.e. do not change their action to manipulate the future behaviour of others.[7] Formally, in this simulation we assume that in each period each player maximizes the probability of making a correct choice, i.e.

$$a_{i,t} = \begin{cases} R \text{ if } p_{i,t} > 0.5 \\ G \text{ if } p_{i,t} < 0.5 \end{cases} \tag{1}$$

where $p_{i,t}$ denotes the probability player i assigns to state R at the beginning of period t. When $p_{i,t} = 0.5$, the player is indifferent and we assume that she randomizes uniformly over the two actions. Each player computes this probability using Bayes rule

[7] This assumption is commonly made in the literature, see e.g., Parikh and Krasucki (1990); Gale and Kariv (2003); Mossel et al. (2014); Harel et al. (2021) and seems plausible given the complexity of the environment.

taking into account the actions of others. We assume common knowledge of rationality, so each player knows that the others are making their decisions according to (1). We thus assume that each player does not only know the signal generating process, but also the strategies of all other players, which is arguably a strong rationality requirement. In Table 2 we report the probabilities of correct actions under these assumptions. We give these simulation results as a benchmark for what Bayesian agents can hope to achieve in this setting. Our results do not hinge on a comparison of the subjects' behavior to this benchmark.

5 Results

We present results from our experiments in the following order. First, we focus on the public data setting, and compare the SIGNALS and ALL treatments (Section 5.1). Second, we study the private data setting (the ACTIONS treatment) and investigate group size effects (Section 5.2). As we go through the analysis, we summarize the main empirical findings as observations.

Data Analysis Approach. Our main analysis considers all ten games in all sessions.[8] We investigate the effect of the treatment on the fraction of correct actions and on the consensus rates using regression analysis. To compare performance across treatments, we estimate a linear regression of the indicator of a correct action by a given participant in a round of a game on a round-specific treatment effect. To compare consensus rates across treatments, we regress the relative size of the majority by round on a round-specific treatment effect.[9]

Throughout our analysis, we classify the collection of agents' signals according to their strength, as given by the difference in the number of signals of one color relative to the other. We define five categories of signal strength: *Very Strong Green*, *Strong Green*, *Weak*, *Strong Red*, and *Very Strong Red*. These correspond, respectively, to the difference between the number of red and green signals being less than -26, in $[-26, -15]$, in $[-14, 14]$, in $[15, 26]$, and above 26. The boundaries between these

[8]We see moderate learning across games within a session. Section D in the Online Appendix shows aggregate results subsetting our data for early (first 5) and late (last 5) games.

[9]Appendix A reports the average treatment effects across rounds as well as treatment effects in early (rounds 2 to 10) and late rounds (rounds 11 to 20). All standard errors are clustered at the session level to account for the inter-dependencies of observations that come from re-matching subjects within a session.

categories correspond to the 10^{th}, 25^{th}, 75^{th} and 90^{th} percentile of the distribution at the end of the game. For the small group treatments we adjust the intervals to match the same percentiles.[10]

In some discussions, we only distinguish between "weak" and "strong" signals by pooling together the *Strong* and *Very Strong* categories for each color. An alternative approach of defining the signal strength based on the *fraction of signals* instead of the difference in the number of signals of each color yields similar results. We discuss these two approaches in detail in Section 5.1.

5.1 The Public Data Setting and the Effect of Redundant Information

In this section we compare the ALL and SIGNALS treatments. Panel (a) in Figure 1 presents our main outcome of interest: how often subjects guess the state correctly in each round of the game, in comparison to the Bayesian benchmark.

In both treatments, the first action that subjects make is a coin toss. This is expected since the this action is made before any signals are observed, so the probability of guessing correctly is one half. As the game progresses and more information arrives, we observe an upward trend in the likelihood of choosing the correct action.

The main finding of this paper is that subjects perform better in the ALL treatment, as compared to the SIGNALS treatment. This happens despite the aforementioned fact that the only difference between the two treatments is the observability of other players' actions, which are informationally redundant. Across rounds, the percentage of correct actions is on average 5% larger in the ALL treatment, as compared to the SIGNALS treatment.[11] In the next section, we explore in depth why this is the case. We note that both treatments significantly under-perform relative to optimal behavior, which means that in both treatments at least some subjects deviate from the Bayesian benchmark.

Panel (b) of Figure 1 measures the polarization of opinions, presenting the evolution of consensus rates. The consensus rate is defined as the relative size of the majority subgroup in a given round, based on members' actions. This rate varies between one half and one, where a consensus rate of one half indicates a maximally polarized group

[10] In section E of the Online Appendix we show that the qualitative results reported in this section are robust to changing the cutoffs of these categories.

[11] See panel (a) of Figure A.1 in Appendix A and columns (1) and (2) in Table A.1 for point estimates and standard errors.

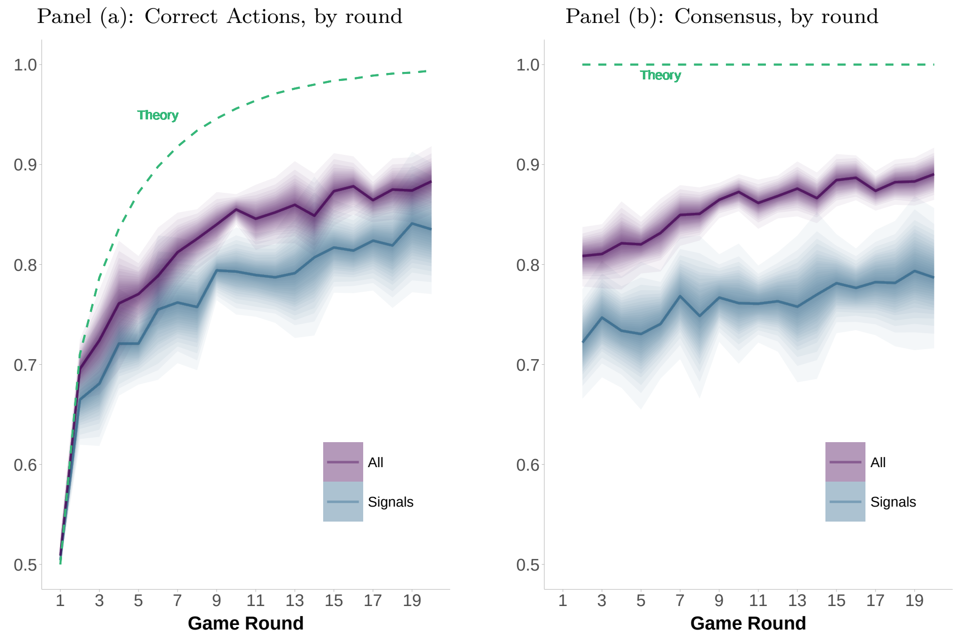


Figure 1: Aggregate statistics in the ALL and SIGNALS treatments

Notes: Panel (a) presents the average frequency of correct actions in each round, averaged across games. Panel (b) depicts the evolution of consensus in each round, i.e., the relative size of the majority, averaged across games. For panel (b) we exclude cases with an equal number of green and red signals. Shaded regions represent 95% confidence intervals from 50% (darkest) to 95% (faintest) probability levels. Confidence intervals are constructed with a variance-covariance matrix clustered by session.

with half of the members choosing each of the two actions, while a consensus rate of one indicates the case where all members choose the same action in a particular round.

Overall, the consensus rates are increasing in both treatments as the game progresses, showing that group members' opinions tend to align the longer they interact with each other. This is to be expected, since actions become more correlated with the state, and hence with each other, as subjects gather more information. Nevertheless, consensus rates are significantly less than one, which would be the theoretical prediction for Bayesian agents.[12] More interestingly, the consensus rates are significantly higher in the ALL treatment than in the SIGNALS treatment: subjects agree more when they see others' actions and raw data, than when they only see raw data. On

[12]Note that Panel (b) excludes cases in which the number of red and green signals is the same, for which theory does not provide a unique prediction on the size of the majority.

average, consensus rates are 10% higher in the ALL treatment than in the SIGNALS treatment.[13] This is consistent with subjects being influenced by the actions of others.

Finally, focusing on cases with strict majorities, we note that the probability that the majority is correct is extremely high and exceeds 90% on average in both the ALL and SIGNALS treatments. Furthermore, it is 4% higher in the ALL treatment, as compared to the SIGNALS treatment.[14]

Observation 1: *People learn faster and develop more unified opinions when they observe each others' signals and actions, compared to observing signals only.*

What Drives These Aggregate Results? To what extent do subjects' actions follow the information contained in the signals they observe? We start with the simplest statistic, i.e., the second round behavior, which, when performed optimally, entails reporting the color of the majority of the eight signals observed in the first round of the game. In both treatments, a large fraction of around 84% of our subjects choose optimally in the second round.

As more and more signals are observed, a Bayesian subject would keep a tally of the signals and report the majority color. Panel (a) in Figure 2 shows the frequency with which subjects chose the color red as a function of the observed difference between the total number of red and green signals. Panel (b) shows the estimated probability of an optimal action, i.e., reporting the color of the majority of signals, as a function of signal strength.

This figure suggests several insights. First, while panel (a) would look like a step function for a Bayesian (the green line), in the data we see a gradual increase in the frequency of choosing red. Thus, subjects respond to signals, and do so in the correct direction, but not perfectly. Second, the mistakes are much more frequent when signals are weak, i.e., when the majority signal color has occurred only slightly more than the other color. Because the expected utility is a function of the difference between the number of red and green signals, this is consistent with a Luce model of behavior, in which choice probabilities are related to the difference in expected payoffs (Luce, 1958). It is also consistent with standard models of perception where close-by states are harder to distinguish. Third, subjects systematically perform better in the

[13] Additional evidence is presented in panel (b) of Figure A.1 and average point estimates and standard errors are in columns (3) and (4) in Table A.1 in Appendix A.

[14] The evidence is presented in Figure A.2 and columns (5) and (6) of Table A.1 in Appendix A.

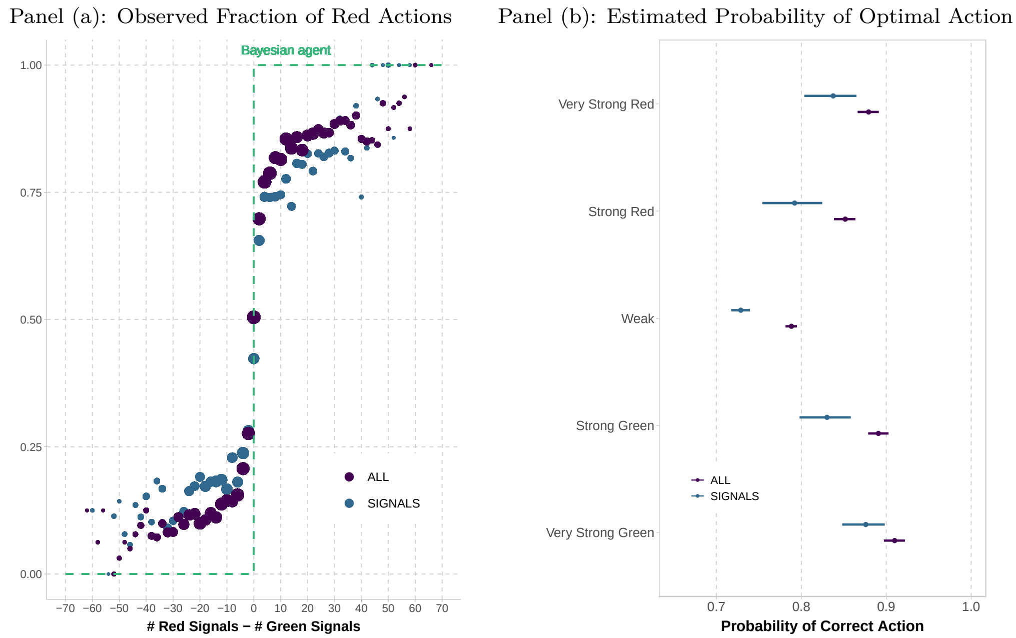


Figure 2: Learning from signals in the SIGNALS and ALL treatments

Notes: Panel (a) depicts the fraction of red actions as a function of the difference between the number of red and green signals. The size of the dot corresponds to the number of observations in each bin. Panel (b) shows the estimated probability of an optimal action (i.e., reporting the color of the majority of signals) as a function of signal strength. We estimate a Bayesian logistic regression of a red action on our measure of signal strength and session random effects. We present the estimated posterior median and 95% confidence intervals. The categories in Panel (b) for signal strength are as defined in Section 5.

ALL treatment across different signal strengths. Particularly, when signals are weak subjects are 6% more likely to choose correctly in the ALL treatment compared to the SIGNALS treatment (73% for the SIGNALS versus 79% for the ALL treatment).[15]

Figure 3 provides a closer look at the behavior in the ALL treatment. To show the usefulness of observing other people's actions, in addition to the raw data, we estimate a Bayesian logistic regression of a subject's action on the share of others' actions (in the previous round), conditional on the different categories of signal strength, as previously defined. To account for interdependencies of subjects within a session, we

[15]For the recovered probabilities on panel (b) of Figure 2, we estimate a Bayesian logistic regression of an indicator of a red action on the signal strength and session random effects. As robustness, Table F.2 in the Online Appendix shows similar estimates under different linear probability models and clustered standard errors by session. We obtain similar results when we account for game, round and participant effects.

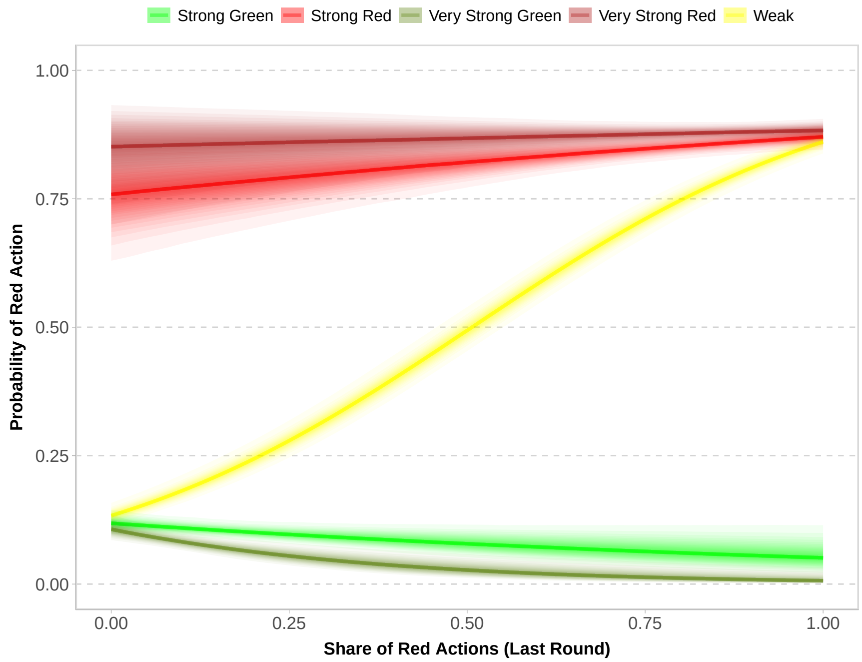


Figure 3: Learning from others' actions in the ALL treatment

Notes: This figure depicts the probability of choosing red as a function of the share of red actions of other group members. The estimates are obtained from a Bayesian logistic regression of subjects' actions on the share of others' actions in the previous round conditional on signal strength and session random effects. Shaded regions represent 95% confidence intervals from 50% (darkest) to 95% (faintest) probability levels.

control for session random effects. We find that when signals are strong or very strong, a subject's actions are barely influenced by those of others: regardless of what others do, subjects tend to follow the majority of signals. This can be seen by mostly flat lines in the bottom and top of Figure 3. In contrast, when the number of signals of each color is comparable, subjects rely heavily on the actions of others, as is indicated by the steep middle line, which shows a strong responsiveness of subjects to others' actions.

An alternative approach, in which one classifies signal strength by the *fraction* of signals instead of the difference in the *number* of signals of each color yields similar results. This alternative approach is inspired by prior experiments, which found that some people rely on sample proportion over sample size in probability judgement tasks (see, e.g., Benjamin, 2019).[16]

Observation 2: *In both the* SIGNALS *and* ALL *treatments, subjects use signals to choose their actions, but do so less well than in the Bayesian optimum. In particular, subjects tend to make more mistakes when the tally of signals is close, i.e, when signals are not very informative. In the* ALL *treatment, subjects partially correct for these mistakes by relying on others' actions when signals are weak, which accelerates learning as compared to the* SIGNALS *treatment.*

Why Does Relying on Others Improve Performance? Individual Level Analysis. Our results so far show that conditioning on the social signal, i.e., on the actions of others, improves performance in the ALL treatment, as compared to the SIGNALS treatment. The question is why? To investigate this question, we calculate a number of statistics describing a subject's behavior in our game and explore its joint distribution in the population. We analyze subjects' behavior by their level of IQ, given our auxiliary IQ measure.

We classify subjects into low-IQ and high-IQ based on their answers to six IQ questions: three matrix reasoning questions, which are similar to Raven's Progressive Matrices, and three 3-D rotation questions (Chapman et al., 2019).[17] The low-IQ subjects are those who answered at most half of the IQ questions correctly and the high-IQ subjects are the remaining group.[18]

We find that our IQ measure, despite its obvious coarseness, correlates with subjects' performance in the main game: high-IQ subjects are approximately 9% more likely to guess the state correctly than the low-IQ ones in both the ALL and

[16] One could ask which of the two models fits our data best, i.e., do people behave as if they condition their actions on the fraction or on the number of signals of each color? Our data slightly favors the former model, but the difference is truly small: The average fraction of correctly predicted guesses is 72% *vs* 71%. Figure E.2 in the Online Appendix re-does Figure 3 using the fractions of signals to compute signal strength and presents the in-sample fit in terms of fraction of correctly predicted guesses from the two alternative approaches.

[17] The questions are drawn from the International Cognitive Ability Resource, a public domain intelligence measure (ICAR; Condon and Revelle (2014)). In each of the first three questions, participants are asked to determine which of the options completes a graphic pattern. In the remaining questions, participants are asked to identify which of the presented drawings of a cube were compatible with another drawing of a cube. In general, these questions have been shown to capture a variation of fluid intelligence in the general population and is becoming one of the popular measures of IQ on par with CRT tests and general knowledge tests.

[18] The average number of correct answers is 3.7 in the ALL treatment and 3.4 in the SIGNALS treatment. This classification delivers roughly similar proportions of low-IQ subjects across treatments: 47% in the ALL treatment and 40% in the SIGNALS treatment.

the SIGNALS treatments ($p < 0.01$).[19] Figure 4 presents the raw fractions of optimal actions by IQ type and signal strength. It shows that (a) high-IQ subjects are more likely to take the optimal action compared with low-IQ subjects irrespective of signal strength and (b) conditional on the IQ type, all participants take the optimal action more often when signals are strong than when they are weak.

Next, we investigate whether there are systematic differences in how low-IQ and high-IQ subjects process the information available to them. We start with the SIGNALS treatment and calculate two statistics for each subject: (i) *responsiveness to weak signals*: the probability that they choose optimally (i.e., with the majority of signals, as a Bayesian would) when signals are weak and (ii) *responsiveness to strong signals*: the probability that they choose optimally when signals are strong. This probability is computed via a Bayesian logistic regression of the subject's action on our measure of signal strength. To compute responsiveness at the individual level, we exploit the variation across rounds and games for a given subject, which allows us to recover measures of responsiveness at the individual level via individual-specific random effects.[20] Panel (a) in Figure 5 presents a scatter plot with each dot corresponding to one participant depicting responsiveness to strong and weak signals. Panel (c) underneath summarizes the same information as the kernel distributions of individual estimates. The estimates of individual responsiveness are given by the posterior median of the probability of an optimal action by the participant.

We perform the same estimation for subjects playing the ALL treatment, but interacted with the social signal measured by the actions of other group members in

[19]Tables G.1 and G.2 in the Online Appendix show the estimates and clustered standard errors of a linear probability model of correct actions on the 6-question IQ measure and the low-IQ/high-IQ indicator, respectively. The relationship between accuracy and IQ is robust to controlling for game and round effects as well as for other subject characteristics.

[20]Specifically, the regression takes the form

$$Pr(v_{it}^{g}) = f\left(\alpha_i + \sum_j \beta_{j_i} strength_j + \gamma_{\text{session}}\right), \ \alpha_i \sim \mathcal{N}(0, \sigma_\alpha), \beta_{j_i} \sim \mathcal{N}(0, \sigma_\beta) \text{ and } \gamma_{\text{session}} \sim \mathcal{N}(0, \sigma_\gamma) \qquad (2)$$

where v_{it}^{g} equals one if participant i guessed optimally in round t of game g and zero otherwise. $strength_j$ takes the value of one whenever the signal strength lies within category $j \in \{Weak, Strong\}$. γ_{session} is a random effect per session. As we are estimating a logistic regression, $f(\cdot) = logit^{-1}(\cdot)$. For a fully Bayesian participant i we expect to see 100% responsiveness for both weak and strong signals.

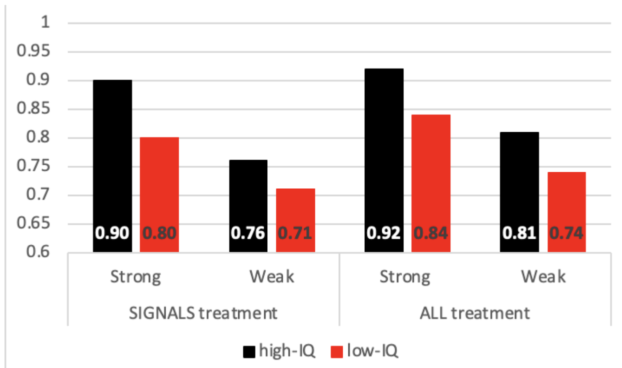


Figure 4: Optimal Actions by Signal Strength and IQ, individual level data

Notes: Each bar represents the fraction of observed optimal actions (i.e., consistent with the majority of signals) across subjects and over rounds by IQ type and signal strength. We exclude cases with an equal number of green and red signals.

the previous round.[21] To isolate the responsiveness to raw signals, we set the number of others who take each action to be equally split between red and green. These results are presented as a scatter plot in Panel (b) and as kernel distributions in Panel (d) of Figure 5.

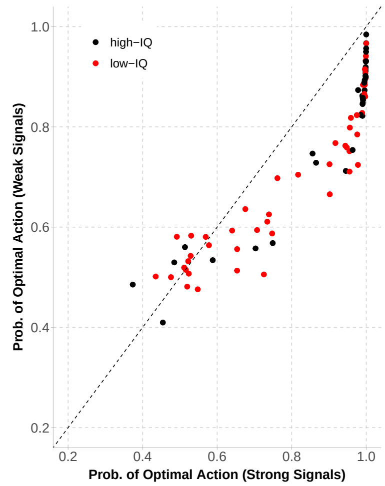
Two patterns are apparent from Figure 5. First, in both treatments, the high-IQ participants are on average much more responsive to raw signals than the low-IQ ones, both when signals are weak and when they are strong. In fact, high-IQ subjects are very close to the Bayesian benchmark when it comes to strong signals, but some of them make a significant number of mistakes with weak signals. These mistakes are, however, less frequent than the mistakes of the low-IQ subjects, which explains why low-IQ subjects perform worse than the high-IQ ones in the SIGNALS treatment.

Second, there is substantial heterogeneity in individual responsiveness to raw signals among participants in both treatments. For instance, about a quarter of high-IQ subjects in the ALL treatment have estimated responsiveness of above 90% for both weak and strong signals, compared to only 7% of low-IQ subjects with similar estimates. This means that the social signal, measured by the average action of all

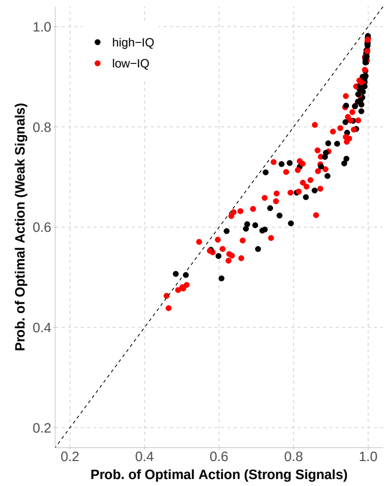
[21] Specifically, the Logit regression takes the form

$$Pr(v_{it}^{g}) = f\left(\alpha_i^{'} + \sum_j \beta_{j_i}^{'} strength_j \times actions_{t-1} + \gamma_{\text{session}}^{'}\right), \ \alpha_i^{'} \sim \mathcal{N}(0, \sigma_{\alpha'}), \beta_{j_i}^{'} \sim \mathcal{N}(0, \sigma_{\beta'}) \text{ and } \gamma_{\text{session}}^{'} \sim \mathcal{N}(0, \sigma_{\gamma}) \quad (3)$$

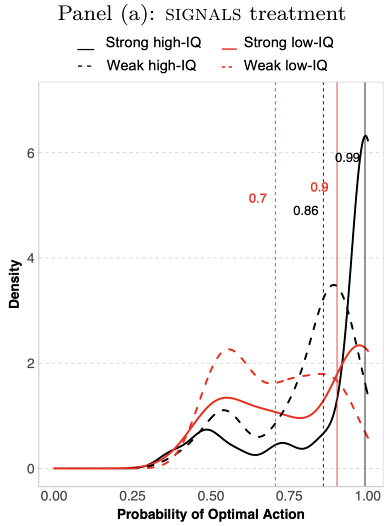
where $actions_{t-1}$ denote the share of others' red actions in the previous round $t-1$.



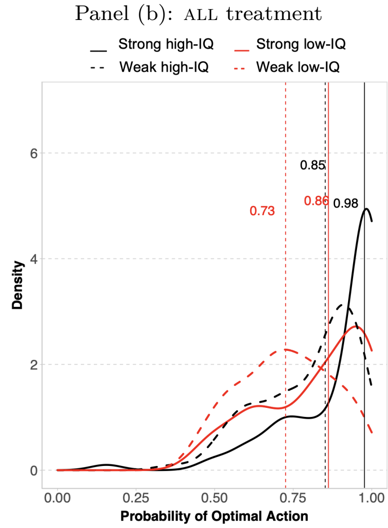
Panel (a): SIGNALS treatment



Panel (b): ALL treatment



Panel (c): SIGNALS treatment



Panel (d): ALL treatment

Figure 5: Responsiveness to signals, individual level data

Notes: Panels (a) and (b) are scatter plots of participants' responsiveness to signals, with horizontal axes depicting responsiveness to strong signals and vertical axes the responsiveness to weak signals. In these scatter plots, each dot represents one participant. Kernel distributions of participants' responsiveness to signals are presented in Panels (c) and (d). These kernels depict weak signals and strong signals separately for each group of participants. The vertical lines and the numbers next to them are median responsiveness for each group. Responsiveness to signals is calculated based on regressions (2) and (3) for the SIGNALS and ALL treatments, respectively. Responsiveness is given by the probability that a participant's action is optimal. For the ALL treatment, we also set the actions of others group members in the previous round to be uninformative, i.e., split equally between green and red.

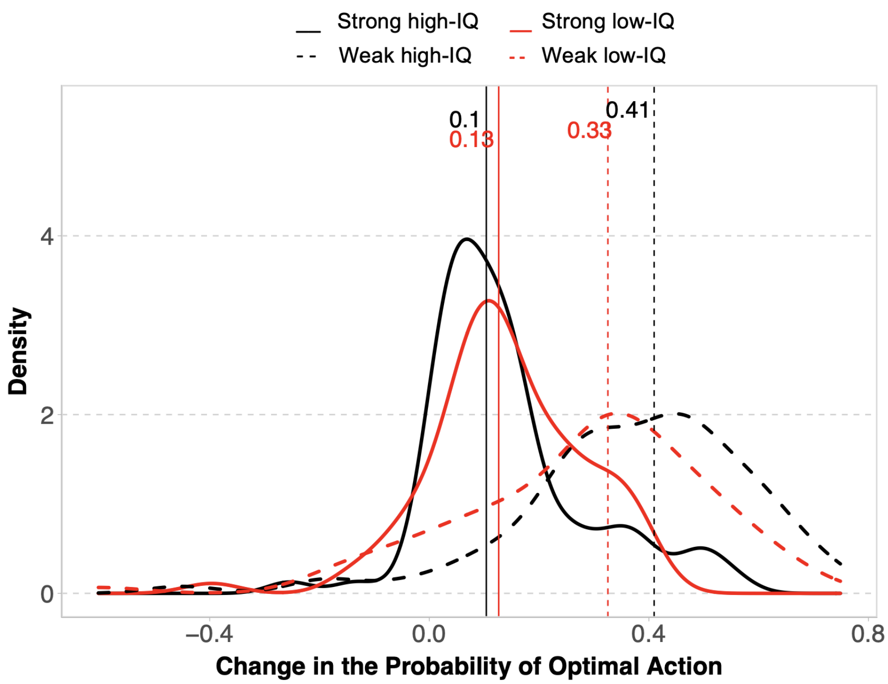


Figure 6: Responsiveness to actions, individual level data

Notes: Kernel distributions of participants' responsiveness to others' actions are presented separately for weak and strong signals. The vertical lines and the numbers next to them depict median responsiveness for each group. Responsiveness to others' actions is measured by the change in the probability of choosing the action of the majority of signals when all versus none of the other group members choose the majority of signals in the last round.

group members in the ALL treatment, is informative since it contains a sizable fraction of correct actions.

Who uses this social signal, and does it affect one's performance? Figure 6 sheds light on this question. For both low-IQ and high-IQ participants, we compute individual responsiveness to social signals as the difference in the probability of choosing optimally when all others agree with the majority of signals minus this probability when all others disagree with the majority of signals.[22] As can be seen in the figure, both groups of participants rarely condition on others' actions when signals are strong, but they do so to a substantial degree when signals are weak. On

[22]E.g., if a subject chooses optimally with probability 69% when all the last round actions that she observes agree with the majority of signals, and chooses optimally with probability 42% when all these actions disagree, then her responsiveness to social signals would be 27%.

average, the high-IQ subjects condition on others to a *larger* extent than the low-IQ ones ($p = 0.004$). Most importantly, given that the social signal is informative as we have argued above, conditioning on it is beneficial for all participants, not just the low-IQ ones. Those low-IQ participants whose responsiveness to others' actions for weak signals is above the median in their group, guesses the state correctly 86% of the time, which is significantly higher than the 70% performance those below the median ($p < 0.001$). The same applies to high-IQ participants, with 91% vs 81% chance of guessing correctly for those above and below the median, respectively ($p = 0.001$). In other words, sufficiently many participants aggregate information well, making the social signal informative and worth conditioning on.

We also study the relationship between participants' IQ types and their strategies, as elicited in the series of open-ended questions at the end of the experiment. We estimate a probabilistic topic model known as Structural Topic Model (STM) (Roberts et al., 2013), which is an unsupervised machine learning algorithm for probabilistic text classification. In essence, STM represents each participant's response by a distribution of latent word clusters or topics. We recover the proportion of both low-IQ and high-IQ subjects' responses spent covering each estimated topic. Appendix B shows a strong consistency between subjects' actions in the experiment and the way they describe their strategies. First, high-IQ participants described using strategies based on own and others' signals to a larger extent than low-IQ players. Second, both low-IQ and high-IQ subjects described using others' actions when making their own choices in similar proportions. Third, the subjects who found the information on others' actions redundant come disproportionately from the group of high-IQ players.

Similarly, we find consistency between subjects' actions in the experiment and their beliefs. In Appendix C we show the results of regressing a subject's beliefs accuracy, given by the mean squared error across belief questions, on our IQ measure, while controlling for treatment effects and subject characteristics. Overall, we find that high-IQ subjects are significantly more accurate than low-IQ ones in forming their beliefs across treatments.

Observation 3: *The high-IQ subjects process the raw signals and perform well, regardless of whether social information is available. The low-IQ subjects do not process well the information contained in the signals, and hence perform badly when social information is not available. When social information is available, it is used by subjects in both groups and boosts the performance of those who use it.*

5.2 The Private Data Setting and the Effect of Group Size

We now turn our attention to the ACTIONS treatment, in which people observe each others' actions but not others' signals. We ask whether they can parse useful information from these choices. This task is complex since actions of group members are inherently correlated and making such inferences requires forming beliefs about how others act. Even under heroic assumptions about common knowledge of rationality among group members, theory is only able to characterize long run outcomes such as the speed of learning, but not how individuals should act in the short term (Harel et al., 2021). Our experimental data is particularly useful in these circumstances as it provides the first step at documenting how people actually behave in this complex situation.

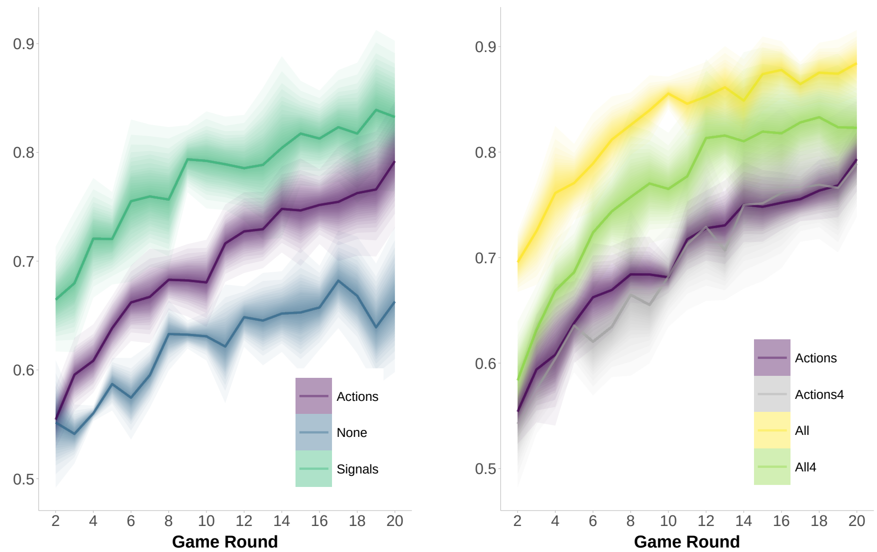
Panel (a) in Figure 7 compares the performance in the ACTIONS treatment with two benchmarks: the NO INFO and the SIGNALS treatments. The former benchmark is the lower bound of learning rates, which is what one expects to happen in the ACTIONS treatment if subjects ignore entirely each others' actions and base their decisions only on the sequence of private signals they receive. The latter benchmark is the upper bound, in which the signals of all members are public; this is what would happen if subjects could perfectly infer the signals of others from their actions.

Panel (a) shows that, despite the complexity of the inference problem in the ACTIONS treatment, people are able to extract useful information from others' actions and learn faster than they would without this information, as the comparison to the NO INFO treatment shows. For instance, notice that subjects chose the correct action with a higher probability in period 13 of the ACTIONS treatment than in period 20 of the NO INFO treatment. Thus, the benefit of observing others' actions eventually exceeds the benefit of the additional 7 signals which are potentially revealed by the agents' second period actions. The same holds more generally if we compare period t of the ACTIONS treatment with period $t + 7$ of the NO INFO treatment. However, if people observe each others' raw signals instead of actions then they learn even faster, as the comparison between ACTIONS and SIGNALS shows.[23] Both of these patterns are qualitatively consistent with theoretical predictions.

[23] Table A.2 and panel (a) of Figure A.3 in Appendix A confirm that the difference in performance of ACTIONS vs NO INFO treatment is significantly higher in later rounds (9.6%) compared with early ones (5.3%) with $p < 0.1$. Table A.2 in Appendix A compares performance of NO INFO vs ACTIONS treatments as well as ACTIONS vs SIGNALS treatments separately for rounds 2 - 20 and rounds 11 - 20. In all pairwise comparisons we obtain significant difference between all treatments.

Figure 7: Frequency of correct actions, by information structure and group size

Panel (a): Information Structures Panel (b): Group Size



Notes: Both panels present the average frequency of correct actions in each treatment per each round, averaged across games. Shaded regions represent confidence intervals from 50% (darkest) to 95% (faintest) probability levels. Confidence intervals are constructed with a variance-covariance matrix clustered by session.

Panel (b) of Figure 7 explores the effect of group size on learning rates in the two information structures: one with private signals, the ACTIONS treatments, and another with public signals, the ALL treatments. Interestingly, larger groups learn faster only when both signals and the actions of others are public (ALL4 versus ALL treatments), but not when group members observe each others' actions only (ACTIONS4 versus ACTIONS treatments).[24] The lack of a group size effect on learning in the ACTIONS treatment is reminiscent of the theoretical result described in Harel et al. (2021), according to which the speed of learning does not change with the group size. This is, however, a quite loose interpretation of the theoretical result, given that the theory refers to the long-run effects at the limit as time tends to infinity. With regards to the level of consensus across group sizes, we find smaller groups to be more aligned than larger groups, especially in early rounds of the game.[25]

[24]Table A.3 and panel (b) of Figure A.3 in Appendix A confirm these results. Specifically, the probability of a correct action is 10% higher in the ALL compared with ALL4 ($p < 0.01$), whereas this difference is not statistically different from zero when comparing ACTIONS4 vs ACTIONS.

[25]See Figure F.1 and Table F.1 in the Online Appendix for details.

Observation 4: *Larger groups learn faster than small groups when people have access to both others' actions and others' signals. However, when people can only rely on others' actions, the group size does not affect the speed of learning.*

What drives these aggregate patterns? In this section, we explore the mechanism behind *Observation 4*. We first study the ACTIONS treatments, in which subjects observe the actions, but not the signals of others. Figure 8 shows how responsive our subjects are to others' actions separately for the ACTIONS4 and ACTIONS treatments.[26]

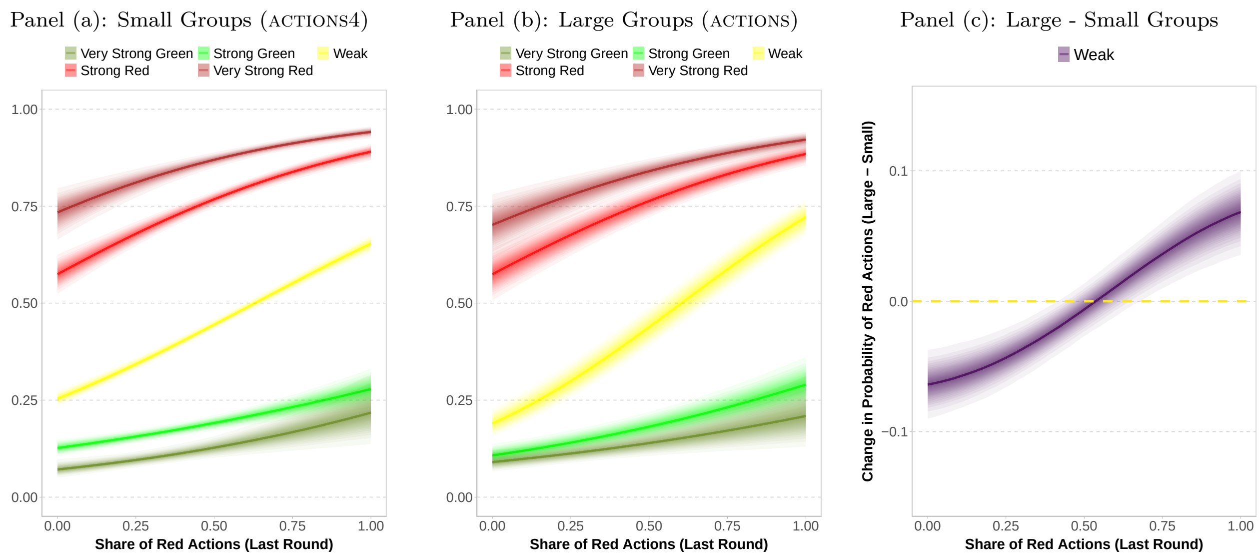


Figure 8: Learning from others' actions in ACTIONS treatments

Notes: Panel (a) depicts the probability of choosing red as a function of the share of red actions of other group members in the ACTIONS4 treatment, obtained from a Bayesian logistic regression of the subject's action on the share of others' actions in the previous round conditional on the difference between red and green signals. Shaded regions represent 95% confidence intervals from 50% (darkest) to 95% (faintest) probability levels. Panel (b) presents the same exercise for the ACTIONS treatment. Panel (c) shows the difference in the probability of a red guess between the ACTIONS and ACTIONS4 treatments for weak signals.

Irrespective of group size, subjects condition on others when signals are weak, as can be seen by the steepness of the *Weak* lines in panels (a) and (b), and by the difference in probability between large and small groups in panel (c). This is qualitatively similar to the behavior of subjects in the ALL treatments.[27] But even

[26] This is the same exercise as we have done for the ALL treatment reported in Figure 3.

[27] See Figure G.1 in the Online Appendix for a comparison between ALL4 and ALL treatments.

when signals are strong, subjects pay considerable attention to others' actions.[28] This behavior shows that participants understand the value of others' actions, since those encode to some extent others' private signals, which are hidden in this treatment.

There are some magnitude differences in how much people rely on others in the small and large groups in the ACTIONS treatment: subjects condition more on others in the larger groups, as can be seen in panel (c) of Figure 8. In the regression analysis that follows we explore whether these differences have a meaningful effect on performance in the two treatments.

Table 3 details the results of four regressions that study the extent to which subjects are influenced by others, as well as the effect that this influence has on performance. In column (1), we study the ACTIONS4 and ACTIONS treatments, and regress the indicator of the event that a subject chooses red on the fraction of others who chose red in the previous round, controlling for our discrete measure of signal strength as well as for game, round and participant fixed effects. We estimate this effect separately for large and small groups. For both group sizes, the effect is substantial: when the rest shift from choosing green to red, a subject's probability of choosing red increases by 26% in small groups, and by 40% in large groups. This difference is significant, which is consistent with Figure 8.

In column (3), we show the estimates from regressing the indicator of matching the state with the fraction of others that match the state, in the ACTIONS4 and ACTIONS treatments, separately. This speaks to the extent that relying on others contributes to a subject's performance. The table shows that there is a small but insignificant difference between the two group sizes: even though (as column (1) shows) subjects rely significantly more on others in larger groups, this reliance does not produce a significant improvement in performance. These results correspond to the classic tension of social learning, namely that relying on the social signal reduces the amount

[28]For instance, in both the ACTIONS and ACTIONS4 treatments, players whose own signals are almost all green (i.e., lie within the *Very Strong Green* category), respond to observing others' changing from a green to a red action by more than doubling their own probability of choosing red (i.e., from 9% to 21%). Individual results corroborate these aggregate results. Figure G.2 in the Online Appendix shows individual responsiveness to signals and others' actions of low-IQ and high-IQ participants in the ACTIONS treatment, using the same classification method as we have done in the ALL and SIGNALS treatments. We find that both low-IQ and high-IQ subjects rely on actions of others both when signals are weak and when they are strong, but more so when signals are weak. Furthermore, when signals are weak, high-IQ subjects on average rely more on the others' actions as compared with low-IQ subjects, which is consistent with their higher ability to learn from others' actions.

Table 3: Magnitude of Rely-on-Others Effect and its Contribution to Performance

	Dependent variable: Action = Red	
	(1)	(2)
	ACTIONS treatment	ALL treatment
Small group: Others' Red Actions (frac.)	0.263***	0.459***
	(0.023)	(0.015)
Large group: Others' Red Actions (frac.)	0.398***	0.700***
	(0.045)	(0.031)
Signal Strength Fixed Effects	Yes	Yes
Game Fixed Effects	Yes	Yes
Game Round Fixed Effects	Yes	Yes
Participant Fixed Effects	Yes	Yes
Obs (Subjects), Small/Large	12,773 (76)/22,916 (136)	13,860 (80)/ 27,752 (152)
Adjusted R^2, Small/Large	0.375/0.324	0.469/0.453
	Dependent variable: Action = State	
	(3)	(4)
	ACTIONS treatment	ALL treatment
Small group: Others' Correct Actions (frac.)	0.190***	0.282***
	(0.025)	(0.039)
Large group: Others' Correct Actions (frac.)	0.245***	0.193***
	(0.043)	(0.050)
Signal Strength Fixed Effects	Yes	Yes
Game Fixed Effects	Yes	Yes
Game Round Fixed Effects	Yes	Yes
Participant Fixed Effects	Yes	Yes
Obs (Subjects), Small/Large	12,773 (76)/22,916 (136)	13,860 (80)/ 27,752 (152)
Adjusted R^2, Small/Large	0.305/0.307	0.346/0.225

Notes: *p<0.1; **p<0.05; ***p<0.01. Clustered standard errors by session in parentheses.

of information revealed by actions, which then in turn degrades the quality of the social signal. Ultimately, in our setting, the result is that larger groups do not perform significantly better than small ones in the ACTIONS treatments.

In columns (2) and (4) we repeat these exercises for the ALL4 and ALL treatments, and draw similar conclusions: subjects rely more on the actions of others in the larger groups, but this reliance does not lead to better performance. In fact, others' actions are more beneficial in smaller groups rather than large ones. Since performance is overall higher in larger groups (see panel (b) of Figure 7), we can conclude that this improvement is not due to reliance on social signals, but simply to the availability of more raw data.

Observation 5: *Subjects respond more to the actions of others in larger groups than they do in smaller groups. But, as individual actions become less informative, this does not result in improved performance in larger groups.*

6 Discussion

The imitation of the actions, opinions and values of others is a natural human tendency. It is an integral part of how infants and adults learn, and of how information is disseminated through society. It can also lead to inefficient outcomes, as has been highlighted by the social learning literature over the past few decades. In our setting, which features repeated interaction between participants and gradual accumulation of public raw data as well as others' opinions, we show that participants imitate each other, and that imitation improves their performance.

This result was not a-priori obvious. Optimal Bayesian behavior does not feature imitation, and behavioral economics does not offer a clear-cut prediction that imitation should improve efficiency, or indeed suggests the opposite. Clearly, it is highly inefficient for all agents to ignore the raw data and mimic each other, and so too much imitation is harmful. Hence the usefulness of imitation lies in heterogeneity: some agents need to base their decisions on the raw data even if others do rely on their peers.

Participants naturally reverted to imitation when signals were "weak" to correct for mistakes, which are common in this case. Grouping participants by IQ results reveals that while high-IQ participants perform better than low-IQ ones irrespective of the information structure and signal strength, both groups rely on the opinions of others when signals are "weak" and both groups benefit from imitation. Imitation and its beneficial effect was robustly observed across information structures and group sizes.

A fruitful avenue for future research is the development of a theoretical approach that explains our findings. The theory of Bayesian agents predicts that actions will be based on the raw public data, and so observing the actions of others should not improve outcomes. Thus a departure from the Bayesian framework is needed to explain our results. A theory whose predictions match our observations will necessarily need to incorporate heterogeneity, as well as some measure of trust in others' actions.

To study imitation we constructed a social learning environment with repeated actions. The experimental literature on such settings is small, and we believe that there is much room for further research. In particular, it would be interesting to understand the robustness of imitation to various interventions, such as the inclusion of a well-informed central information source, or of ideological agents whose actions are independent of the data, or a more general structure of heterogeneity in information quality of different agents. These questions are important in an age in which the dissemination of uncertain but critical information to the public is a major societal challenge.

A Additional Aggregate Results

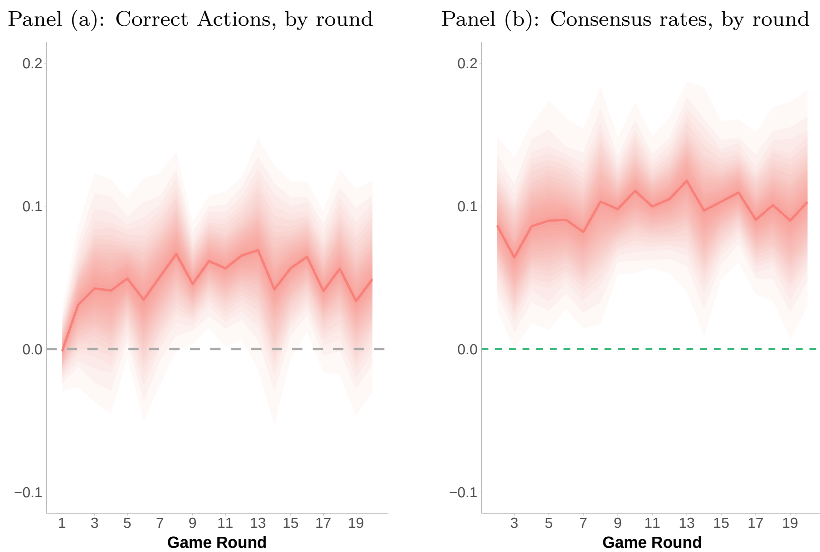


Figure A.1: Differences between the ALL and SIGNALS treatments

Notes: Panel (a) presents the average difference in the frequency of correct actions in each treatment in each round, averaged across games. Panel (b) depicts the difference in consensus rates in each round. For panel (b) we exclude cases with equal number of green and red signals. Shaded regions represent confidence intervals from 50% (darkest) to 95% (faintest) probability levels. Confidence intervals are constructed with a variance-covariance matrix clustered by session.

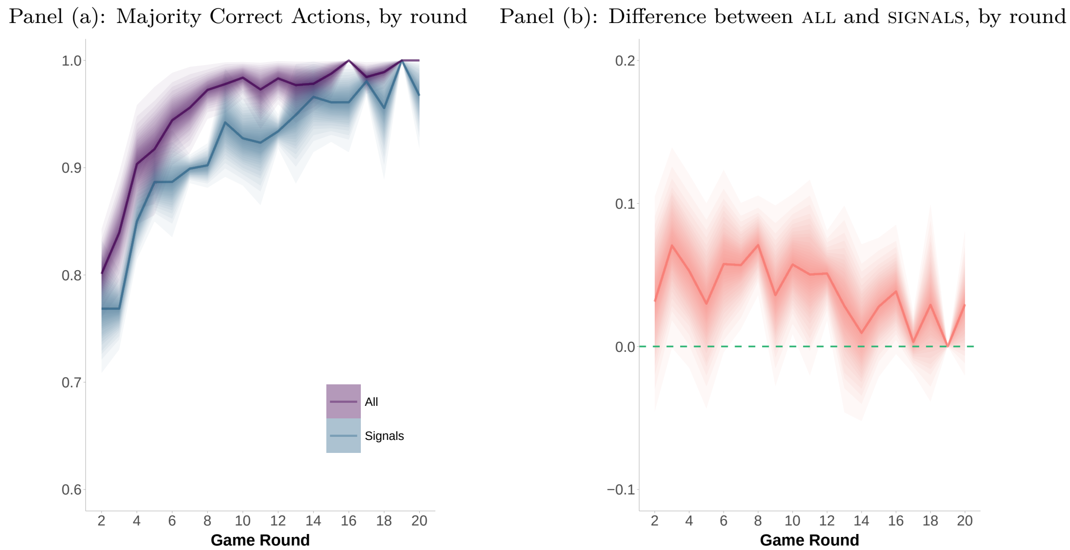


Figure A.2: How often is the majority correct? ALL versus SIGNALS treatments

Notes: Panel (a) presents the frequency of correct actions by the majority in each treatment per round, averaged across games. Panel (b) presents the average difference in the frequency of correct actions by the majority between the two treatments. We exclude cases with an equal number of green and red signals. Shaded regions represent confidence intervals from 50% (darkest) to 95% (faintest) probability levels. Confidence intervals are constructed with a variance-covariance matrix clustered by session.

Table A.1: Treatment Effects for SIGNALS and ALL treatments

	Dependent variable:					
	Correct Actions		Consensus Rate		Majority is Correct	
	(1)	(2)	(3)	(4)	(5)	(6)
ALL (Baseline)	0.702***	0.701***	0.812***	0.809***	0.803***	0.808***
	(0.012)	(0.013)	(0.009)	(0.011)	(0.013)	(0.014)
SIGNALS (Effect)	−0.051*	−0.047*	−0.095***	−0.089***	−0.038**	−0.051***
	(0.028)	(0.028)	(0.028)	(0.029)	(0.016)	(0.016)
SIGNALS (Effect) × Late Rounds		−0.007		−0.011		0.026*
		(0.017)		(0.012)		(0.015)
Game Round Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	44,460	44,460	5,499	5,499	5,499	5,499
Adjusted R^2	0.021	0.021	0.135	0.135	0.062	0.062

Notes: **p<0.05; ***p<0.01. Clustered standard errors by session in parentheses. Late rounds are 11 through 20.

Table A.2: Treatment Effects for NO INFO, ACTIONS, and SIGNALS

	Dependent variable:	
	Correct Actions	
	(1)	(2)
ACTIONS (Baseline)	0.583***	0.572***
	(0.017)	(0.015)
NO INFO (Effect)	−0.076***	−0.053***
	(0.019)	(0.016)
NO INFO (Effect) × Late Rounds		−0.043*
		(0.026)
SIGNALS (Treatment)	0.078***	0.096***
	(0.029)	(0.028)
SIGNALS (Treatment) × Late Rounds		−0.034**
		(0.013)
Game Round Fixed Effects	Yes	Yes
Observations	56,430	56,430
Adjusted R^2	0.028	0.028

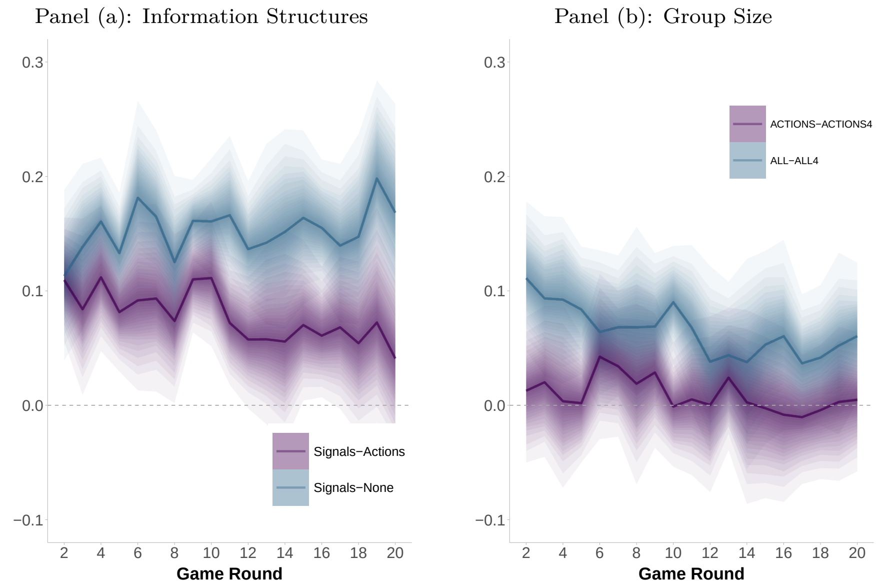
Notes: **p<0.05; ***p<0.01. Clustered standard errors by session in parentheses. Late rounds are 11 through 20.

Table A.3: Treatment Effects for Group Size

	Dependent variable:			
	Correct Actions			
	(1)	(2)	(3)	(4)
ALL4 (Baseline)	0.614***	0.603***		
	(0.017)	(0.016)		
ACTIONS4 (Baseline)			0.546***	0.539***
			(0.016)	(0.012)
ALL (Effect)	0.065***	0.082***		
	(0.023)	(0.022)		
ALL (Effect) × Late Rounds		−0.033		
		(0.031)		
ACTIONS (Effect)			0.008	0.019
			(0.022)	(0.020)
ACTIONS (Effect) × Late Rounds				−0.021
				(0.024)
Game Round Fixed Effects	Yes	Yes	Yes	Yes
Observations	44,080	44,080	40,090	40,090
Adjusted R^2	0.029	0.029	0.020	0.020

Notes: **p<0.05; ***p<0.01. Clustered standard errors by session in parentheses. Late rounds are 11 through 20.

Figure A.3: Frequency of correct actions, by information structure and group size



Notes: Both panels present the average frequency of correct actions in each treatment per each round, averaged across games. Shaded regions represent confidence intervals from 50% (darkest) to 95% (faintest) probability levels. Confidence intervals are constructed with a variance-covariance matrix clustered by session.

B Open-ended Questions

We quantify the answers to the open-ended questions designed to elicit participants' strategies throughout the game and analyze their relationship with participants' characteristics, specifically players' IQ.

For each participant in the ALL4 and ALL treatments, we examine the three open-ended survey questions asked after all rounds have been played:

(i). What strategy did you use in the game (if any)? Please elaborate.

(ii). Did you look at the balls drawn for other players in your group? Did you find them useful/not useful? Please elaborate.

(iii). Did you look at the bets made by other players in your group? Did you find them useful/not useful? Please elaborate.

We implement a machine learning algorithm for a probabilistic topic model known as the structural topic model (STM) (Roberts et al., 2013). Under this framework, a *topic* is defined as a probability distribution over words and a participant's response in our data is modeled as a distribution over topics. Thus, each participant's response in the data can belong to multiple topics with a probability distribution estimated from the data. We are interested in the proportion of an answer spent covering each estimated topic.

Each participant's response to question $q \in \{(i), (ii), (iii)\}$, r_q, has its own distribution over topics, θ_{r_q}. We label this parameter topic prevalence and interpret it as the proportion of each topic $k = 1, \ldots, K$ in response r_q. We can think of each topic k as drawn from a multinomial distribution with parameter θ_{r_q}. Conditional on the topic selected, word $w_{r_q,n}$ included in response r_q is drawn from a multinomial distribution over the vocabulary $n = 1, \ldots, N$ with parameter $\beta_{k,n}$. This is a probability vector over the N words in the vocabulary.

In probabilistic topic models such as the STM, the number of topics K needs to be selected a priori. In doing so, there is a trade-off between interpretability (consistent with a lower K) and goodness-of-fit (consistent with a higher K). We favor the former and choose the low value of $K = 4$ for the three open-ended questions. We find that $K = 4$ gives us topics that are very straightforward to interpret and with a higher semantic coherence (i.e., the most probable words in a given topic co-occur together) than models with more topics.

The structural topic model allows for the inclusion of participants' characteristics to inform the topic prevalence. Specifically, $\theta_{r_q} \sim \text{LogisticNormal}(X_{r_q}\gamma, \Sigma)$, where X_d is a vector of participant characteristics. In addition to our binary measure of IQ, we include several participant characteristics: `female`, which is an indicator variable that takes the value of one if the participant identifies as female. `stem`, which is an indicator variable that takes the value of one if the participant's major is STEM. `overconfidence` measures the extent of a participant's over-estimation of her IQ, which is given by the difference between the number of questions a participant believes she solved correctly and the actual number of correct answers. `risk` is measured by the number of points invested in a risky asset as specified in a risky investment task solved at the end of the game.

The topic model for each open-ended question is estimated using a variational Expectation-Maximization (EM) algorithm, as implemented in the `stm` package in `R` (Roberts et al., 2013). Prior to estimation, we pre-processed the raw responses using standard conventions: we stem words (i.e., reduce words to their root form), drop punctuation, as well as common stop-words, and remove words that were used less than 0.5% over all responses.

Figure B.1 shows the labels of three estimated topics for the strategy question along with the actual responses that are estimated to be highly associated with each topic. The topics we labeled *Other Strategies* encompass participants who either choose actions at random, ignoring signals and actions of others, or used heuristics that deviate from Bayesian updating. The topic we labeled *Signals + Actions of Others* encompasses answers where participants emphasized using both signals and others' actions to make their choices. The topic labeled *Signals* describes answers where participants said making their choices based on the aggregate number of green and red balls drawn throughout a game. Figures B.3 and B.5 show the labels and response examples for questions *(ii)* and *(iii)*, respectively. For these questions, the estimated topics capture the range from negative to positive reliance on signals and others' actions, respectively.

Figure B.2 shows for each estimated topic in the strategy question *(i)*, the relationship between our measure of IQ and the estimates of topic prevalence, θ_{r_q}. We capture each topic with a word cloud of the top words associated with it. Figures B.4 and B.6 present these estimates for questions *(ii)* and *(iii)*, respectively.

We find that participants' description of their strategies is consistent with their actual behavior in the game according to their IQ type. In particular, high-IQ participants, who appear closer to Bayesian behavior in the game (see Figure 5), described using strategies based on signals to a larger extent than low-IQ ones (14% larger with $p < 0.001$, see panel (a) in Figure B.2) who, in turn, relied significantly more on heuristics which, in many cases, disregarded raw signals as well as actions of others (see panel (b) in Figure B.2). Moreover, when directly asked "Did you look at the balls drawn for other players in your group? Did you find them useful/not useful?", low-IQ participants were more skeptic about the information provided by other players' signals, compared to high-IQ players (5% more with $p < 0.001$)

In terms of explicitly looking at the actions of others, the topic prevalence from question *(i)*, which we labeled "Signals + Actions of Others" shows that high-IQ participants describe other players' actions as useful in a similar proportion to low-IQ ones (37% for high-IQ participants versus 42% for low-IQ ones ($p = 0.055$), see panel (c) in Figure B.2). Moreover, the analysis of the answers to "'Did you look at the bets made by other players in your group? Did you find them useful/not useful?" shows that high-IQ participants describe other players' actions as useful similarly to low-IQ ones (36% *vs* 29% with $p = 0.07$, see panel (a) of Figure B.6), while low-IQ players described the social signal as sometimes/somewhat useful to a larger extent than high-IQ ones (4% and 30% more ($p < 0.05$) for topics "Sometimes/Somewhat Useful", respectively). Overall, these answers are consistent with a similar responsiveness to others' actions by IQ types found in the game data (see Figure 6). At the same time, more than two thirds of those participants who find the information in others' actions redundant come from the group of high-IQ players (see panel (b) in Figure B.6), which is consistent with the behavior of a subset of high-IQ participants who take decisions consistent with Bayesian updating.

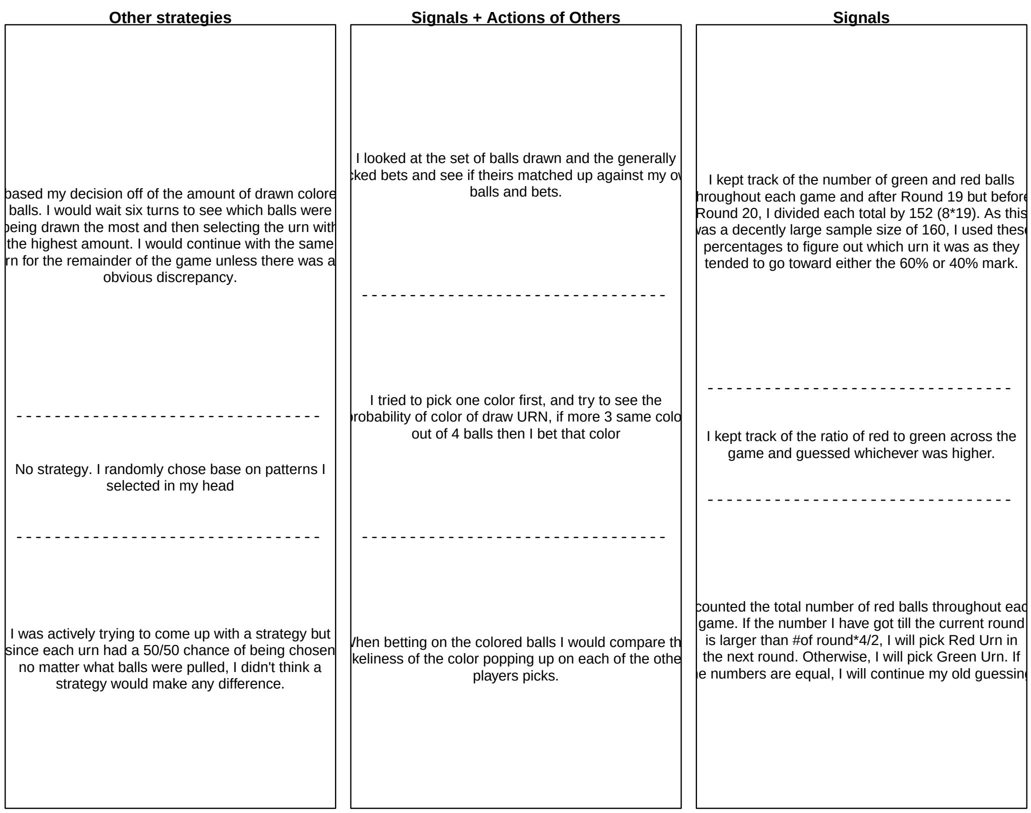


Figure B.1: Example Responses to "What strategy did you use in the game (if any)?" by topic (ALL4 and ALL treatments)

Notes: Each panel presents the three most highly associated open-ended responses with each of the four estimated topics for the question "What strategy did you used in the game (if any)?"

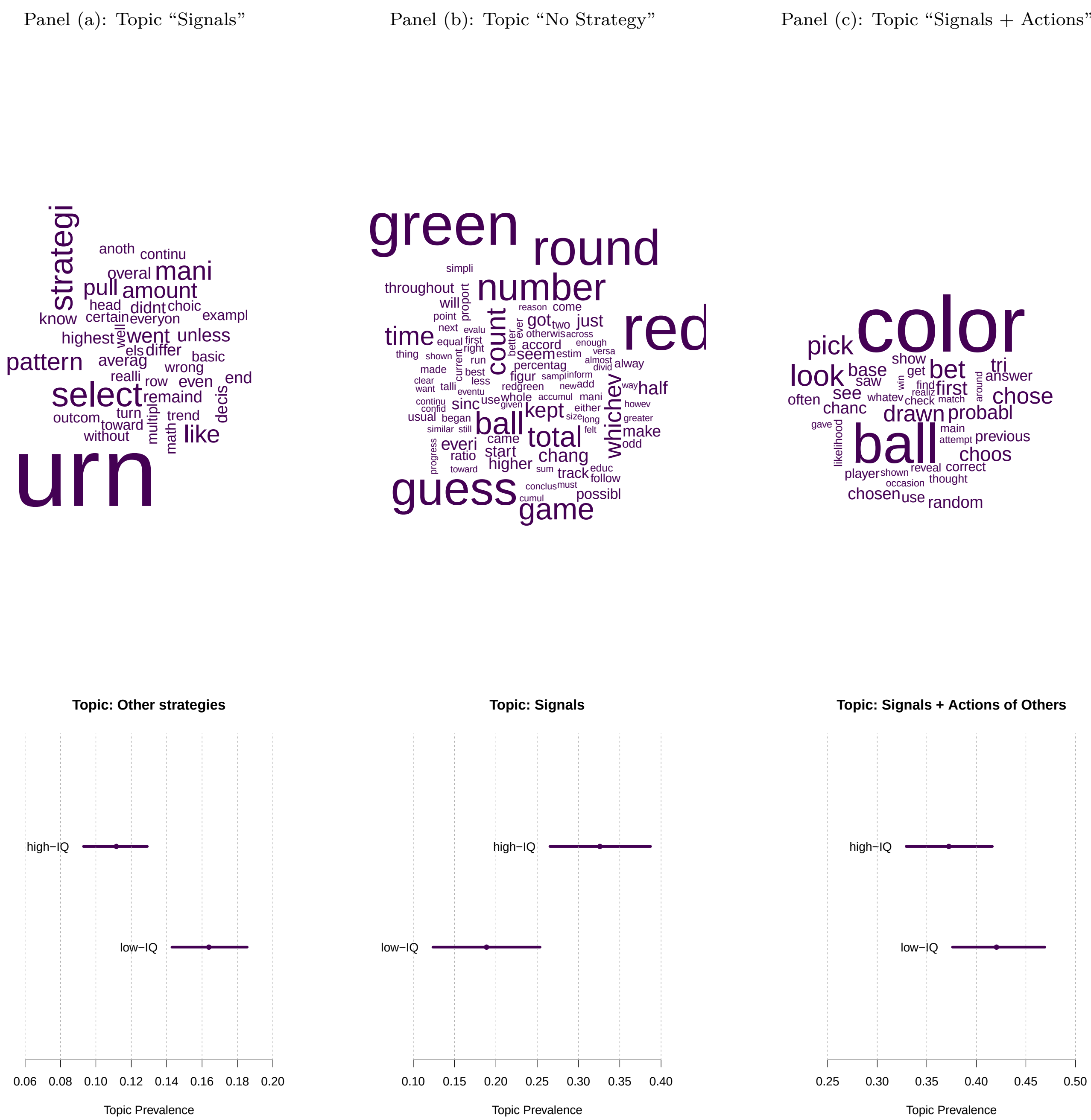


Figure B.2: Responses to “What strategy did you use in the game (if any)?” of low-IQ and high-IQ Participants (ALL4 and ALL treatments)

Notes: The upper row of each panel presents the word cloud of the top 100 words with the highest likelihood of being drawn from each topic, with the size of the word representing the magnitude of this likelihood. The lower row of each panel presents the estimated topic prevalence for the subgroups of low-IQ and high-IQ participants with 95% confidence intervals.

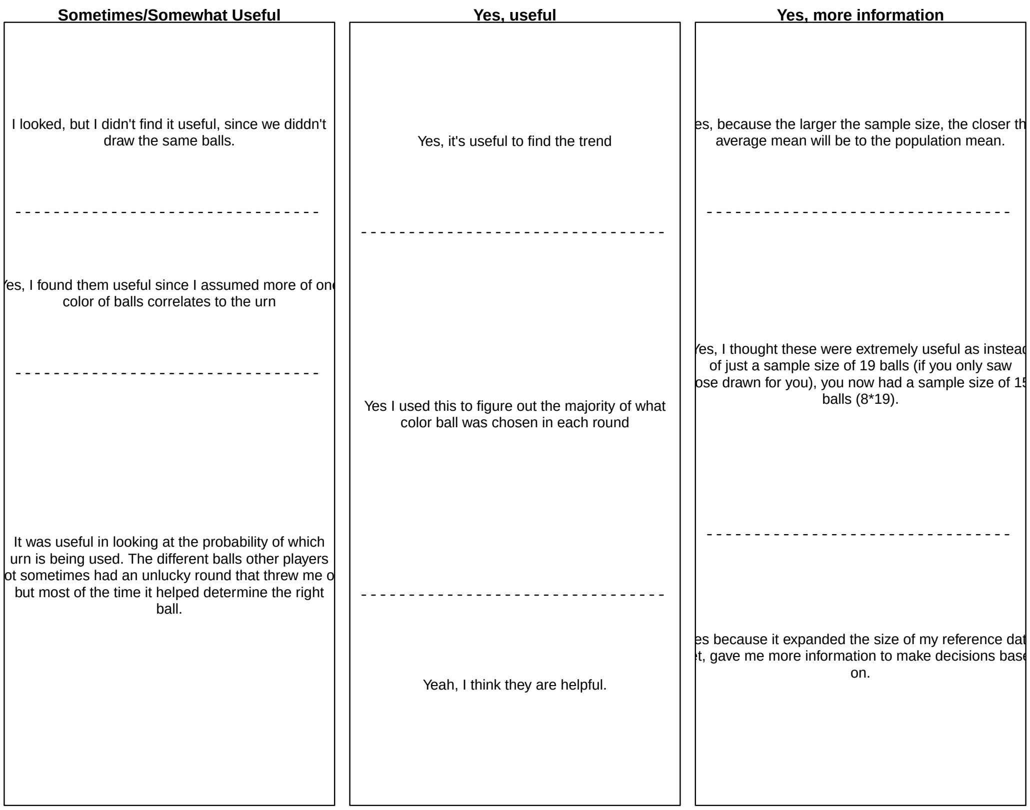


Figure B.3: Example Responses to "Did you look at the balls drawn for other players in your group?" by topic (ALL4 and ALL treatments)

Notes: Each panel presents the three most highly associated open-ended responses with each of the four estimated topics for the question "Did you look at the balls drawn for other players in your group? Did you find them useful/not useful? Please elaborate."

Panel (a): Topic “Sometimes/Somewhat Useful” Panel (b): Yes, more information

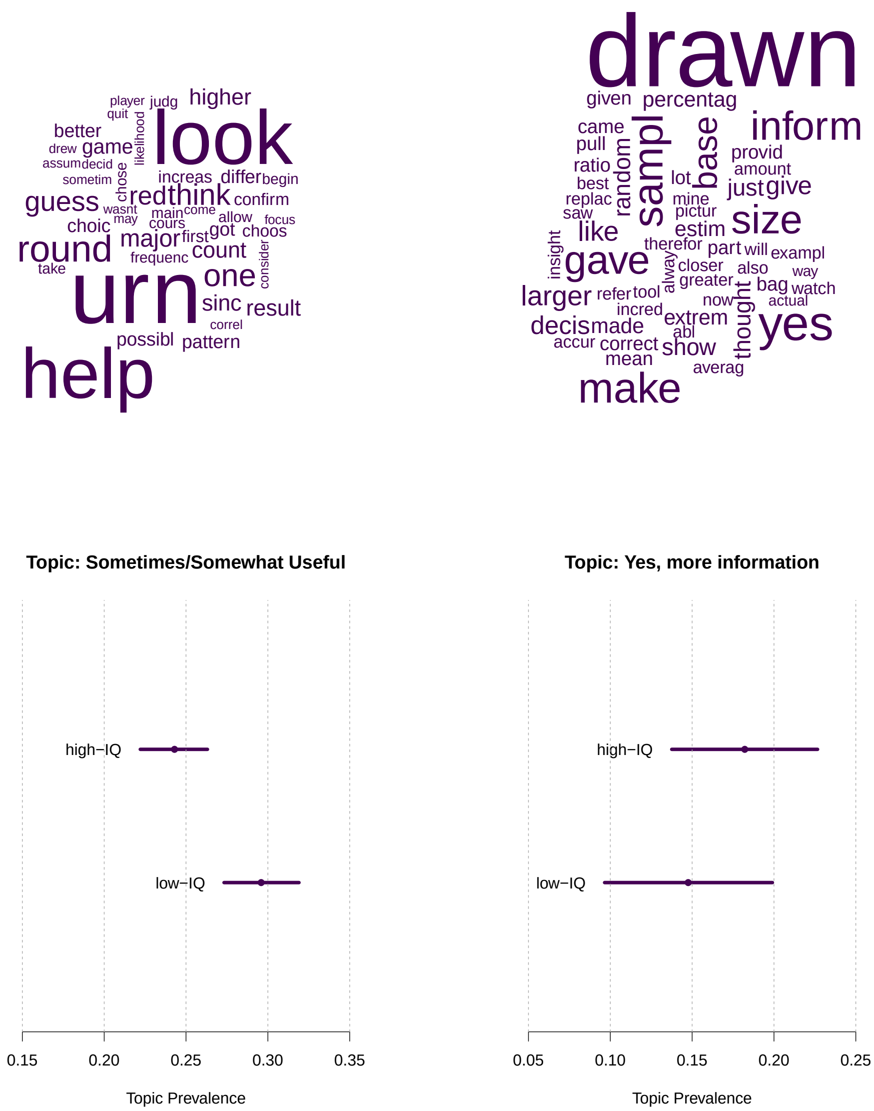


Figure B.4: Responses to “Did you look at the balls drawn for other players in your group?” of low-IQ and high-IQ Participants (ALL4 and ALL treatments)

Notes: The upper row of each panel presents the word cloud of the top 100 words with the highest likelihood of being drawn from each topic, with the size of the word representing the magnitude of this likelihood. The lower row of each panel presents the estimated topic prevalence for the subgroups of low-IQ and high-IQ participants with 95% confidence intervals.

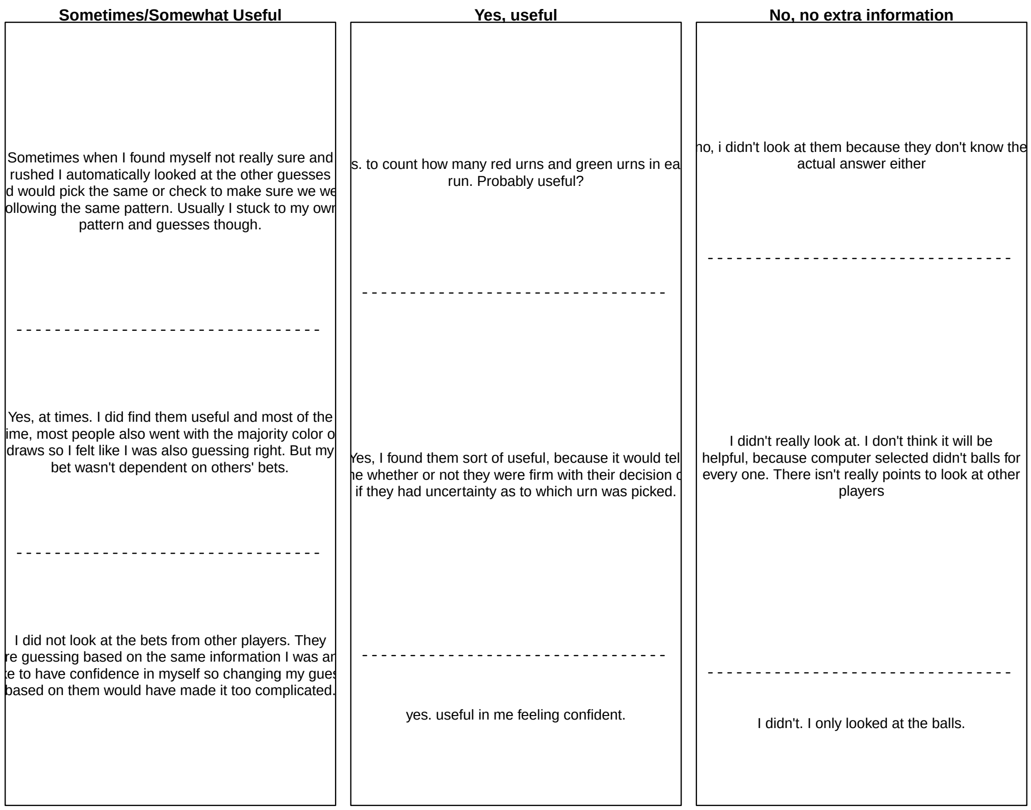
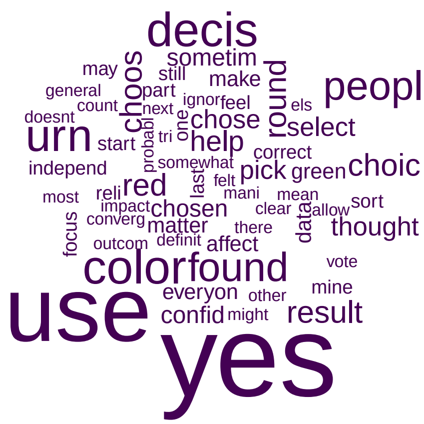
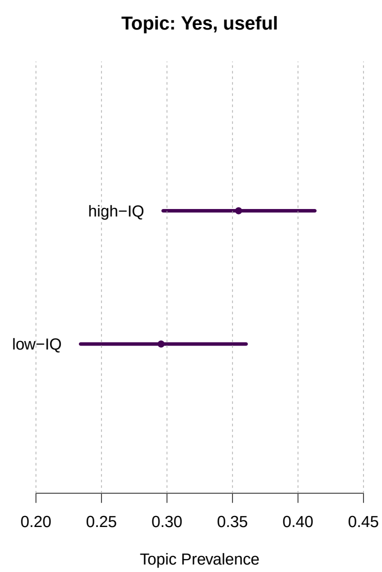


Figure B.5: Example Responses to "'Did you look at the bets made by other players in your group?" by topic (ALL4 and ALL treatments)

Notes: Each panel presents the three most highly associated open-ended responses with each of the four estimated topics for the question "Did you look at the bets made by other players in your group? Did you find them useful/not useful? Please elaborate."

Panel (a): Topic "Yes, useful"





Panel (b): No, no extra information

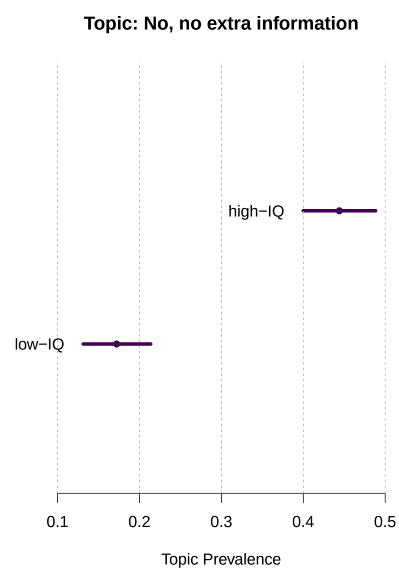


Figure B.6: Responses to "Did you look at the bets made by other players in your group?" of low-IQ and high-IQ Participants (ALL4 and ALL treatments)

Notes: The upper row of each panel presents the word cloud of the top 100 words with the highest likelihood of being drawn from each topic, with the size of the word representing the magnitude of this likelihood. The lower row of each panel presents the estimated topic prevalence for the subgroups of low-IQ and high-IQ participants with 95% confidence intervals.

C Participants' Beliefs

We analyze the precision of subjects' incentivized beliefs, as elicited at the end of the experiment (see section B.3 for details). For the analysis, we pool all participants' answers across treatments.[29] In the beliefs section, each participant provides an estimate of the fraction of correct actions (in the last round of the game) in both her own treatment and other treatments. We combine these answers and define participant i's precision as her mean squared error across all treatments: $MSE_i = \sum_{treat}(y_{treat} - \hat{y}_{i,treat})^2$, where y_{treat} denotes the actual fraction of correct actions in treatment $treat \in \{\textsc{none},\textsc{actions4},\textsc{actions},\textsc{signals},\textsc{all4},\textsc{all}\}$ and $\hat{y}_{i,treat}$ denotes participant i's estimate about the fraction of correct actions in treatment $treat$. Both actual values and beliefs are discretized in 10 equally spaced bins from 0% to 100%, consistent with subjects' answers.

To assess the difference between low-IQ and high-IQ subjects on their beliefs' precision, we estimate a regression of the form:

$$MSE_i = \delta_{treat,i} + \beta_1 \cdot \texttt{low-IQ} + X_i'\gamma + \epsilon_i,$$

where `low-IQ` is a dummy variable that takes a value of one if a subject is low-IQ according to our IQ auxiliary measure and 0 if she is high-IQ. $\delta_{treat,i}$ is a fixed effect of the treatment assigned to subject i. $\mathbf{X}_i$ is a vector of participant i's characteristics as defined in Appendix B including: `female`, `stem`, `overconfidence` and `risk`. ϵ_i denotes the error term with a variance-covariance matrix clustered by session.

The results of estimating the difference between low-IQ and high-IQ subjects on their beliefs' precision is presented in Table C.1 and visually in Figure C.1. Across all specifications, we find that low-IQ subjects are significantly less precise than their high-IQ counterparts. In particular, controlling for both treatment effects and participant characteristics (column (3) of Table C.1), these effect translates into low-IQ participants miscalculating the actual fraction of correct actions by more than 10% (i.e., $\sqrt{1.58} = 1.26$ bins of 10% increments each).

[29] We do not have data for subjects in the first session of each treatment, because it was used to collect data and calibrate subjects' payments in subsequent sessions.

Table C.1: Precision of Beliefs (MSE) by IQ Type (ALL treatments)

	Dependent variable:		
	MSE		
	(1)	(2)	(3)
Intercept	5.281***	3.428***	3.770***
	(0.414)	(0.283)	(0.818)
`low-IQ`	1.097**	1.088**	1.580***
	(0.537)	(0.545)	(0.602)
Treatment Fixed Effects	No	Yes	Yes
Participant Covariates	No	No	Yes
Observations	476	476	476
Adjusted R^2	0.005	0.012	0.013

Notes: *p<0.1; **p<0.05; ***p<0.01. Clustered standard errors by session in parentheses

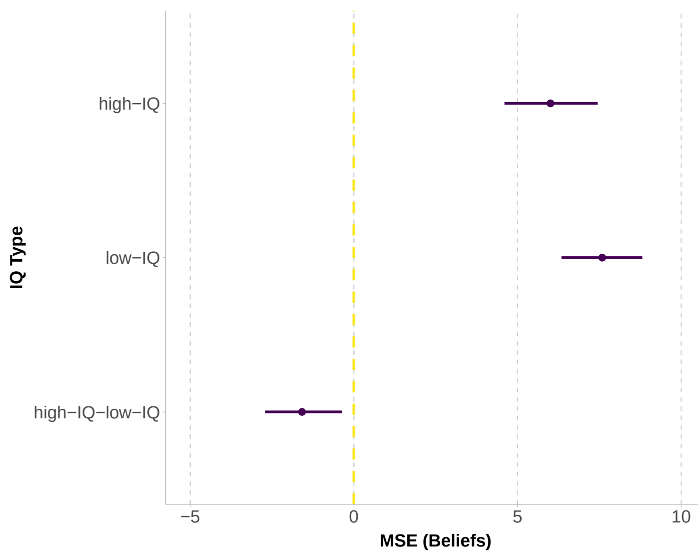


Figure C.1: Beliefs' Accuracy and IQ Types

Notes: The figure shows predicted mean squared errors for the beliefs of low-IQ and high-IQ participants along with 95% confidence intervals. For these estimated effects we use the estimated coefficients from column (3) in Table C.1. We include as controls: `female`, `stem`, `overconfidence` and `risk` and set them to their median values in the data.

References

L. R. Anderson and C. A. Holt. Information cascades in the laboratory. *The American economic review*, pages 847–862, 1997.

M. Angrisani, A. Guarino, P. Jehiel, and T. Kitagawa. Herd behavior in a laboratory financial market. *American Economic Review*, 95:1427–1443, 2005.

M. Angrisani, A. Guarino, P. Jehiel, and T. Kitagawa. Information redundancy neglect versus overconfidence: a social learning experiment. *American Economic Journal: Microeconomics*, 13(3):163–97, 2021.

A. V. Banerjee. A simple model of herd behavior. *The Quarterly Journal of Economics*, pages 797–817, 1992.

D. J. Benjamin. Errors in probabilistic reasoning and judgment biases. *Handbook of Behavioral Economics, edited by Doug Bernheim, Stefano DellaVigna, and David Laibson*, 2019.

S. Bikhchandani, D. Hirshleifer, and I. Welch. A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy*, pages 992–1026, 1992.

S. Bikhchandani, D. Hirshleifer, O. Tamuz, and I. Welch. Information cascades and social learning. Technical report, National Bureau of Economic Research, 2021.

A. Caplin, M. Dean, and D. Martin. Search and satisficing. *American Economic Review*, pages 2899–2922, 2011.

B. Çelen and S. Kariv. Distinguishing informational cascades from herd behavior in the laboratory. *American Economic Review*, 94(3):484–498, 2004.

A. Chandrasekhar, H. Larreguy, and J. Xandri. Testing models of social learning on networks: Evidence from two experiments. *Econometrica*, pages 1–32, 2020.

J. Chapman, M. Dean, P. Ortoleva, E. Snowberg, and C. Camerer. Econographics. *NBER working paper #24931*, 2019.

D. L. Chen, M. Schonger, and C. Wickens. otree—an open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, 9:88–97, 2016.

S. Choi, D. Gale, and S. Kariv. Social learning in networks: a quantal response equilibrium analysis of experimental data. *Review of Economic Design*, pages 135–157, 2012.

D. M. Condon and W. Revelle. The international cognitive ability resource: Development and initial validation of a public-domain measure. *Intelligence*, 43:52–64, 2014.

K. Dasaratha and K. He. An experiment on network density and sequential learning. *Games and Economic Behavior*, 128:182–192, 2021.

R. De Filippis, A. Guarino, P. Jehiel, and T. Kitagawa. Non-bayesian updating in a social learning experiment. *Journal of Economic Theory*, 199:105188, 2022.

J. Duffy, E. Hopkins, T. Kornienko, and M. Ma. Information choice in a social learning experiment. *Games and Economic Behavior*, 118:295–315, 2019.

J. Duffy, E. Hopkins, and T. Kornienko. Lone wolf or herd animal? information choice and learning from others. *working paper*, 2020.

B. Enke and F. Zimmermann. Correlation neglect in belief formation. *Review of Economic Studies*, pages 313–332, 2019.

P. Evdokimov and U. Garfagnini. Individual vs. social learning: An experiment. *working paper*, 2020.

E. Eyster, M. Rabin, and G. Weizsacker. An experiment on social mislearning. *Available at SSRN 2704746*, 2018.

D. Gale and S. Kariv. Bayesian learning in social networks. *Games and Economic Behavior*, 45(2):329–346, 2003.

U. Gneezy and J. Potters. An experiment on risk taking and evaluation periods. *Quarterly Journal of Economics*, pages 631–645, 1997.

J. K. Goeree, T. Palfrey, B. Rogers, and R. McKelvey. Self-correcting information cascades. *Review of Economic Studies*, 74(3):733–762, 2007.

V. Grimm and F. Mengel. Experiments on belief formation in networks. *Journal of the European Economic Association*, 18(1):49–82, 2020.

M. Harel, E. Mossel, P. Strack, and O. Tamuz. Rational groupthink. *The Quarterly Journal of Economics*, 136(1):621–668, 2021.

W. Huang, P. Strack, and O. Tamuz. Learning in repeated interactions on networks. *arXiv preprint arXiv:2112.14265*, 2021.

A. Hung and C. Plott. Information cascades: Replication and an extension to majority rule and conformity-rewarding institutions. *American Economic Review*, 91(5): 1508–1520, 2001.

D. Kubler and G. Weiszacker. Limited depth of reasoning and failure of cascade formation in the laboratory. *Review of Economic Studies*, 71(2):425–441, 2004.

R. D. Luce. A probabilistic theory of utility. *Econometrica: Journal of the Econometric Society*, pages 193–224, 1958.

E. Mossel, A. Sly, and O. Tamuz. Asymptotic learning on bayesian social networks. *Probability Theory and Related Fields*, 158(1-2):127–157, 2014.

M. Mueller-Frank and C. Neri. A general model of boundedly rational observational learning: Theory and evidence. *working paper*, 2015.

R. Parikh and P. Krasucki. Communication, consensus, and knowledge. *Journal of Economic Theory*, 52(1):178–189, 1990.

M. E. Roberts, B. M. Stewart, D. Tingley, and E. M. Airoldi. The structural topic model and applied social science. In *Advances in neural information processing systems workshop on topic models: computation, application, and evaluation*, volume 4, pages 1–20. Harrahs and Harveys, Lake Tahoe, 2013.

B. Scheibehenne and R. Greifeneder. Can there ever be too many options? a meta-analytic review of choice overload. *Journal of Consumer Research*, pages 409–425, 2010.

L. Smith and P. Sørensen. Pathological outcomes of observational learning. *Econometrica*, 68(2):371–398, 2000.

X. Vives. How fast do rational agents learn? *The Review of Economic Studies*, 60(2): 329–347, 1993.

G. Weizsacker. Do we follow others when we should? a simple test of rational expectations. *American Economic Review*, pages 2340–236, 2010.

A. Ziegelmeyer, F. Koessler, B. J., and E. Winter. Fragility of information cascades: An experimental study using elicited beliefs. *Experimental Economics*, 13(2):121–145, 2010.

A. Ziegelmeyer, C. March, and S. Krugel. Do we follow others when we should? a simple test of rational expectations: Comment. *Experimental Economics*, 103(6): 2633–42, 2013.