Title
Essays on Belief-Updating and Decision-Making in Financial Markets

Author
Wang, Zhaoqi

Publication Date
2023

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
SANTA CRUZ

ESSAYS ON BELIEF-UPDATING AND DECISION-MAKING IN FINANCIAL MARKETS

A dissertation submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

ECONOMICS

by

Zhaoqi Wang

March 2023

The Dissertation of Zhaoqi Wang is approved:

Professor Daniel Friedman, Chair

Professor Kristian Lopez Vargas

Professor Natalia Lazzati

Peter Biehl
Vice Provost and Dean of Graduate Studies

Table of Contents

List of Figures

List of Tables

Abstract

Essays on Belief-Updating and Decision-Making in Financial Markets

by

Zhaoqi Wang

This dissertation contains three essays broadly related to financial markets, with an emphasis on decision-makers' belief-updating and decision-making.

Chapter 1 studies subjects' belief-updating when they face an uncertain event accompanied by two independent signals in the laboratory. The "Average model" is introduced and compared with other important models for the goodness off fit. At the individual level, the results are mixed. Some subjects behave close to one of the model's predictions, while some behave close to another model's predictions. The average model outperforms the Bayesian model at the aggregate level in predicting subjects' posterior beliefs. No clear evidence indicates that subjects' posterior beliefs converge to the Average or Bayesian model's predictions over time.

Chapter 2 studies the impact of motivated beliefs, the phenomenon that people believe what they want to believe, on market performance in a laboratory market for a state-contingent common value asset. Motivated beliefs are induced so that traders have polarized preferences over the states. The main findings are that (i) these induced motivated beliefs do not have a significant impact on overall market efficiency, but (ii) they do impact traders' final asset holdings and belief updating processes, and (iii) the induced polarization persists after receiving private signals and trading in the market. Other findings suggest that a more intense financial stake might improve mar-

ket efficiency. The induced ego-relevant motivation is significantly stronger than the homegrown motivation to believe in higher payoffs.

In Chapter 3, joint work with Daniel Friedman and Thomas Bowen, we study the impact of traders' overconfidence on market performance. How does trader overconfidence (judgemental or self-enhancement) affect their performance in asset markets, and overall market quality? Conversely, how does market participation affect traders' overconfidence? To address such questions, we build a laboratory asset market in which human participants receive private information of varying precision and then trade an asset that pays a single state-contingent dividend. Among other results, we find that greater trader overconfidence can improve price efficiency in some environments, but not in the most realistic environment with experienced traders and ambiguous mixed information precision. In that environment, overconfidence reduces trader profits. We detect no substantial impact of market exposure on trader overconfidence.

Acknowledgments

First and foremost, I am extremely grateful to my advisor, Daniel Friedman, for his invaluable advice and continuous support during my Ph.D. journey. Dan's immense knowledge and experience have inspired me every moment of the journey. I wouldn't be here today without his patient guidance in overcoming research challenges and constant encouragement to strengthen my morale. I would like to thank Kristian Lopez Vargas and Natalia Lazzati for their kind help and support that made me a better economist. Without them, this dissertation wouldn't have been possible. I am also grateful to Alan Spearot for his advice and support during my master's journey, which is the start of my wonderful overseas studies in the US.

Next, I would like to thank other faculty members at UCSC who offered constructive feedback on my research. I am grateful to George Bulman, Dominic Massaro, Justin Marion, Ajay Shenoy, Nirvikar Singh, Gerelt Tserenjigmid, and Dong Wei. I would also like to thank Susan Leach, Sandra Reebie, and other UCSC staff for their administrative support throughout the Ph.D. program.

My appreciation also goes out to my inspiring colleagues and coauthors at UCSC: Prateek Arora, Thomas Bowen, Yanshuo Chen, Dongwan Choo, Jijian Fan, Weinan Gong, Mai Hakamada, Naresh Kumar, Jianan Liu, Rongchen Liu, Yilin Li, Liwei Liu, Weicheng Lyu, Ke Ma, David Sungho Park, Akatsuki Sukeda, Brett Williams, Guanghong Xu, Vivian Juehui Zheng, and Shuchen Zhao. They are all my friends. I enjoyed every moment with them in the past few years. Every discussion and chat with them improves my knowledge and being a better person. I am grateful for all their support.

Lastly, I couldn't be more grateful to my parents, Hejian Wang and Luping Yang, for their love and encouragement throughout my life. Home is always the warmest harbor of life. Their unconditional support, whether mental, financial, or emotional, made everything possible. I can't thank them enough.

Chapter 1

Bayesian or Average Updater? Evidence from Lab Experiments

1.1 Introduction

The question about how do people update their beliefs have been widely studied in economics, psychology, and many other social science disciplines. According to Anderson (1981)'s information integration theory, after receiving physical stimuli, people process those stimuli by valuation functions into their psychological values. Then the psychological values are combined by the integration function into an implicit response. Then, through response function, this implicit response is externalized to the observable response.

Physical stimuli and psychological values could take different forms. When investers are making decisions, one of the most common stimuli and values are uncertain events and investers' subjective probabilities. When study uncertainty and probabilities, knowing Bayes rule is a must. Given prior knowledge and conditions that might be related to the events, Bayes rule is known to be the rule that we “should” use to process the information and update precise belief. However, this normative model often fails to predict people’s actual behavior. Field and lab experiments have shown that people systematically depart from Bayesian reasoning (Kahneman and Tversky (1972), Croson and Sundali (2005), Miller and Sanjurjo (2021)). Therefore, finding out the actual valuation and integration processes are very important and have received a lot of attention in the literature.

Depending on the situations and specific tasks, people’s actual behavior differs from Bayes’ predictions in different ways. This paper aims to analyze the situation where people need to integrate two signals regarding an uncertain event. Facing such an event, people will take both signals simultaneously into account if they are perfect Bayesian. However, this paper’s Average model hypothesizes that people would update their belief by treating the two signals separately in a Bayesian way and then take the average of the two to form their final posterior belief.

There are many examples in our life that we might update our beliefs in such an average way. One recent example is about the coronavirus. Back in February 2020, individual investors might be one of the most impacted groups in China due to the nationwide lockdown. News updates the ongoing war against the virus every second and also updates people’s beliefs. People who live in Beijing, for instance, observe two important signals from the media, which says although the situation in Wuhan is still severe (signal 1); in Beijing, however, the number of confirmed/suspected infection is increasing at a decreased rate (signal 2). While signal 1 provides a more negative information, signal 2 is relatively more moderate. Combing those two signals, people think that the turning point has come, many business men and investors started to share their opinions online and looking forward to have their business back to normal. However, with the help of expertise and the correct model of the world, specialists concluded that we were not at the turning point yet and appeal people to keep staying at home for another couple of weeks to prevent the virus spread further. From the above example, we could see that normal people's beliefs deviate from that of professionals'; their posterior belief is more positive since the second signal mitigates the severity of the situation, to which the first signal brings. Such belief updating process is close to what the Average model describes. Therefore, having a better understanding of how people update their beliefs is of crucial importance since it could provide us insight of how people would make their investment decisions when facing uncertain events.

The main justifications for this Average model are: first, the Average model captures representative bias better than perfect Bayesian (Tversky and Kahneman (1974)), and second, this model requires less complication to process the information.

To explore this Average updating rule, this paper will report a lab experiment, where there will be a lottery that takes the value of either 100 or 0 with equal probability. Under each value, there will be an urn that contains the same number of balls with two different colors (black and white) and different proportions. Balls that are drawn from the urn corresponds to the true value of the lottery are signals observed by the subjects. The subjects' task is to update and report their belief of the probability that the lottery worth 100 after receiving new signals. Subjects' posterior beliefs are elicited by using a lottery to lottery exchange method (Karni (2009)) that does not depend on assumptions about risk aversion (A Becker–DeGroot–Marschak method that elicits people's willingness to accept was used in pilots, and it turns out that Karni (2009)'s method worked best). The results show that, at the individual level, while some subjects' behaviors are well explained by one of the tested models, most of the subjects' behaviors cannot be simply categorized into one of the models' predictions. Regressions of elicited data on predicted probabilities show us that the Average models' predictions predict subjects' beliefs better than Bayes model. A couple of other models, including Linear in the Ratio model, Weighted Probability Function, and Bayesian Multiplication, were also tested; and there is no clear evidence that one of the models significantly outperforms others. In addition, there is no clear evidence to show that subjects' beliefs converge to either Bayesian or Average models' predictions over time.

The rest part of this paper is organized as follows. Section 1.2 shows previous works on non-Bayesian updating. Section 1.3 explains the theory of the model. Section 1.4 is the experimental design. Section 1.5 shows experimental results and section 1.6 concludes.

1.2 Relevant Literature

Studying and modeling the way people update their posterior belief have received a lot of attention in the literature. Instead of being consistent and coherent, people's judgment is systematically erroneous due to the law of small numbers (Tversky and Kahneman (1971)), central tendency bias (Crosetto et al. (2020)), overconfidence and optimism (Heger and Papageorge (2018)) or they rely on many heuristic principles such as (1) representativeness, (2) availability of instances and (3) adjustment from an anchor (Tversky and Kahneman (1974)). Unlike perfect Bayesian, people's posterior belief is more close to an inverse-S-shaped weighting function (Tversky and Kahneman (1992), Tversky and Kahneman (1979), Wu and Gonzalez (1996), Camerer and Ho (1994)), and the shape of the probability weighting function is determined by two properties: Discriminability and Attractiveness (Gonzalez and Wu (1999), Prelec et al. (1998)).

There are many heuristics that people could suffer while updating their beliefs in different situations. For example, the amount of information that financial markets provide is often too overwhelming to process; thus, people might use "rule of thumb" to treat markets to have only two states (earnings are mean-reverting and trend) (Barberis et al. (1998)). When facing multiple events, decision makers might assess the degree to which A is similar to B when asked to evaluate the conditional probability of A given B (Zhao (2016)). With access to a sample of the population, the law of small numbers would cause people to make a biased conclusion (Rabin (2002)). But when the number of observations increases to infinity, the ambiguity associated with decision makers' act disappears (Marinacci (2002)). With the access of prior, people might exhibit confirmatory bias that causes them to misreads signals as supporting their current hypothesis and thus become overconfident (Rabin and Schrag (1999)). When people have multiple priors and one of the priors that people initially assign zero probability happened, this will bring people into a situation where Bayes won't help, and they will look over the prior of the priors and readjust their posterior belief (Ortoleva (2012)). While many works consider non-Bayesian updating is not an optimal way of processing information;

Zhang (2013), however, shows that under an evolutionary reproductive process, non-Bayesian updating is the optimal strategy for risk averse agents. In addition, works like Hoffart et al. (2019), pointed out that different situations could have different influence on triggering Bayesian updating.

The situation that this paper focuses on is when an event with uncertainty provides two signals for people to observe. There have been works that tried to study people's behavior when receiving new information successively. Wilde and Li (2019)'s task is to elicit people's belief of the distribution of initial priors and posteriors. The result is that many people are imperfect Bayesian, and about one-sixth adopt non-Bayesian updating. Epstein (2006) pointed out that people have the temptation to overreact to new signals so that if a good signal is realized, people treat this signal as an even better signal about the future states than they had thought before the realization. Epstein et al. (2008) extends the above situation from three periods to infinite periods. In addition, Epstein et al. (2010) studied the case where people over-react or under-react to newly coming signals and proved that those under-react to new signals would eventually forecast accurately. In contrast, those who over-react will forecast accurately with a positive probability. Although the above works suggest alternatives that could explain people's actual behavior, their model suffers from some shortcomings. Below are some heuristics that the above works do not address and they might affect those models' predictions.

Experimental work such as Tversky and Kahneman (1974), Grether (1980), Grether (1992), Grether (1978) and Holt and Smith (2009), provided evidence to show people cannot combine previous beliefs and new information in a Bayesian way. Subjects in their experiments exhibit representative bias, which probabilities are evaluated by the degree to which A is representative of B, that is, by the degree to which A resembles B. Their data were explained well by this heuristic behaviors. Anderson (1981)'s information integration theory provides evidence to show that, although integration of information is undoubtedly a complex process, it often obeys a simple adding-type rule in some situations. Although Massaro and Friedman (1990)'s result shows adding-type model's prediction is non-optimal, experimental evidence from Anderson (1981) illustrates that subjects obeyed an adding-type rule. Observed data support parallelism theorem, which holds true only when people obey adding rule.

The Average model in this paper assumes people take each signal independently and separately and then take the average of the two beliefs to form their final posterior. This Average model can not only reduce the difficulty of processing the information but also be able to better capture representative bias. Experimental data will be collected to test my hypothesis. To achieve my goal, I need to elicit subjects' posterior belief in my experiment, and I adapt the probability elicitation method in Karni (2009), which will be explained in the next section, to directly elicit subjects' posterior probability.

1.3 Theory

1.3.1 Model

When facing two physical stimuli, people might process those two signals separately to obtain two psychological values and then take the average of the two values to form their final belief. Such Average rule has received experimental support (Anderson (1981)) such as ranking personality descriptions and pattern recognition.

Given the prior and signal 1 (s_1), Bayes rule (BU) says the posterior should be $BU(prior, s_1)$. Then, after receiving signal 2 (s_2), Bayes rule says the posterior should be $BU(BU(prior, s_1), s_2)$. However, given prior, s_1 and s_2, Average model says people will update their posterior in the following way:

$$\frac{1}{2}BU(prior, s_1) + \frac{1}{2}BU(prior, s_2)$$

Average model says people might process each signal separately in a Bayesian way and then take the average of the two to form their final posterior.

Our daily life is too complicated to allow us to have a formal model of the environment. However, we can design a lab environment, which is described below, that could capture the most essential characteristics of reality and have subjects perform in the lab environment and see how they behave.

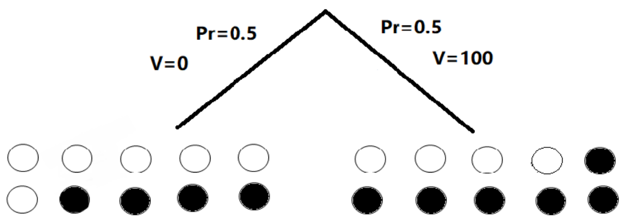


Figure 1.1: Prior and Urn Content for the Hypothetical Lottery

In this experiment, there is a hypothetical lottery that takes values of either 0 or 100 experimental points with equal probability. As figure 1.1 (this figure is on the user interface as public information) shows, subjects know the two values (v=0 or 100) for this lottery and the prior probability (Pr=0.5). After the computer randomly tossing a coin, the true value of the lottery (true state of the world) is determined. Subjects do not know the true value of the lottery, and it is their task to provide their belief of the probability that the lottery worth 100. They will receive two signals, which are balls drawn from the urn corresponding to the true value of the lottery. As shown in figure 1.1, if the lottery worth 100 (resp. 0) points, the computer will use the urn on the right (resp. left) hand side only to draw balls with replacement as the private signal to help the subjects guessing which state is more likely. The lottery would be more likely to worth 100 points if the private signal contains more black balls, since the urn on the right side has more black balls (6 out of 10).

In signal 1 (also called draw 1), the computer will randomly draw $n_1 = 15$ balls in the urn with replacement. Within the n_1 balls drawn from the urn, k_1 of them are black, and m_1 of them are white. Given only signal 1, the Bayesian posterior is:

$$Pr(V = 100|k_1, m_1) = BU(V = 100|k_1, m_1) = \frac{0.6^{k_1} * 0.4^{m_1} * 0.5}{0.6^{k_1} * 0.4^{m_1} * 0.5 + 0.4^{k_1} * 0.6^{m_1} * 0.5}$$

In signal 2 (draw 2), the computer will randomly draw $n_2 = 10$ balls in the same urn with replacement. Within the n_2 balls drawn from the urn, k_2 of them are black, and m_2 of them are white. Similarly, given signal 2, the Bayesian posterior is:

$$Pr(V = 100|k_2, m_2) = BU(V = 100|k_2, m_2) = \frac{0.6^{k_2} * 0.4^{m_2} * 0.5}{0.6^{k_2} * 0.4^{m_2} * 0.5 + 0.4^{k_2} * 0.6^{m_2} * 0.5}$$

Given the above information delivered by the two signals, we want to know how would players update their posterior after observing both signals. If the subject is perfect Bayesian, she should update her posterior based on Bayes Rule:

$$\text{Bayesian Update (BU)} = \frac{A}{A+B} \tag{1.1}$$

Where,

$$A = (0.6^{k_1} * 0.4^{n_1-k_1})(0.6^{k_2} * 0.4^{n_2-k_2})$$

$$B = (0.4^{k_1} * 0.6^{n_1-k_1})(0.4^{k_2} * 0.6^{n_2-k_2})$$

The above Bayes equation, however, is beyond most people's ability to calculate due to the complexity. Also, people may suffer from representative bias; that is, if one of the draws resembles one of the urns while another one resembles another urn,[1] this might weaken the informativeness of each draw, and thus, people's posterior might lie between those two posteriors. Therefore, an average equation can help resolving those two problems, and the Average model is:

$$\text{Average Update (AU)} = \frac{1}{2}p_1 + \frac{1}{2}p_2 \tag{1.2}$$

[1]Here, the resemblance is not in an absolute sense like many other previous works. It does not mean the signal should look exactly like the parent populations in terms of the proportion of black balls in the drawn signal. One signal is considered resemble to one of the urn in terms of black balls' relative proportion compared with another signal.

Where,

$$p_1 = BU(V = 100|k_1, m_1), \text{ and } p_2 = BU(V = 100|k_2, m_2) \tag{1.3}$$

An example that could help us to see the difference between BU and AU is the following: suppose the first signal provides 10 black balls and 5 white balls, while the second signal provides 6 black balls and 4 white balls. We'll have $p_1 = 88.4\%$ and $p_2 = 69.2\%$ by plugging in the numbers into equation (3). This will gives us Bayesian posterior of 94.5% by equation (1), while the Average posterior of 78.8% by equation (2). This means that, similar to the coronavirus example in the introduction section, the second signal supports the true value of the lottery worth 100 less strongly relative to the first signal, and thus subjects might reduce their posterior belief and ended up with 78.8%, which is what Average model predicts.

1.3.2 Elicitation

The elicitation method (Karni (2009)) used in this paper directly elicits subjects' posterior probability that the lottery worth 100, and the method does not depend on subjects' risk attitude. In baseline cases where subjects don't observe any signal, they will be asked to enter their posterior belief of the chances out of 100 that the lottery worth 100. The same question will be asked again when subjects receive signal 1 and also signal 2. To earn more experimental points in this game, subjects' best strategy is to truthfully report their beliefs. So, this eliccitation method is incentive compatible. The explanation for the incentive compatibility is provided in the appendix A.4.5.

1.4 Experimental Design

1.4.1 Parameter Set A Sessions and Parameter Set B Sessions

The experimental design in this paper is as what figure 1.1 shows, I use urns that have black/white ratio of 6:4 in the "good" urn and 4:6 in the "bad" urn. Also, the first signal draws 15 balls, and the second signal draws 10 balls, both with replacement. Subjects are informed how many of the balls are black/white in both of the signals.

To check that the results I got from the above setting are not due to the specific black/white ratio and number of draws I use, I need to change the ratio and draws for some sessions to show that the results remain stable. Therefore, I denote the above settings as parameter set A (Parm-A). Then I change some of the parameters to have parameter set B (Parm-B) sessions.

For parm-B sessions, the black/white ratio in the "good" urn is changed to 7:3, and 3:7 in the "bad" urn. Also, the first draw will draw 5 balls with replacement, and the second draw will draw 3 balls with replacement. Such change is not arbitrary; the change is made in a way so that the Shannon's Entropy in the two types of sessions are similar.[2] (Shannon's Entropy is a measurement of uncertainty and keep it similar in the two types of sessions is important because, in this case, the change of the results, if any, is not due to the change of the uncertainty.) The following introduction of experimental design is based on parm-A sessions' design; the experimental design for parm-B sessions is the same as that of parm-A sessions, the instructions are provided in appendix A.4.

For each session, there are three decision pages for the user interface (UI there-

[2]Proof of the Shannon's Entropy is provided in the appendix at section A.3.1

after). The first decision page asks subjects to report their beliefs truthfully about the probability that the lottery worth 100 without providing any signal. The second decision page asks the subjects the same question providing signal 1. The third decision page asks the subjects the same question again providing signal 2. The three decision pages are under the same state of the world, the computer determines the true state of the world at the first decision page, then the two signals are drawn from the corresponding urn to show the subjects in the second and third decision page. Then subjects could go to the result page for this round showing their payment for each decision page and also the payment for this round. After all 50 rounds finished, subjects will go to the final result page, which shows them the final payment for participating in this session.

I ran five sessions for both parm-A and parm-B sessions,[3] and I have 34 subjects for parm-A sessions and 43 subjects for parm-B sessions. In each round, subjects enter their belief about the chances out of 100 that the lottery worth 100 (a) without observing any signal, (b) after observing signal 1 and (c) after observing signal 2. When a new round begins, the computer tosses the coin again to determine the new value of the lottery and thus the signals in the new round might be different. Subjects play 50 rounds in total. To focus on subjects' updating rules and avoid the effect from learning, the experiment does not provide subjects the true value of the lottery for each round. But the payoff for each round is provided to each subject at the end of each round. The experiment is programmed in oTree platform (Chen et al. (2016)). Each subject is randomly assigned to the computers in UCSC Leeps lab, and they cannot communicate with each other.

[3]Session timelines are: Parm-A: 10/16/2019, 10/23/2019, 11/20/2019, 01/14/2019, 01/21/2019. Parm-B: 02/11/2020, 02/17/2020 (two sessions), 02/18/2020, 02/19/2020.

1.4.2 Demonstration and the User Interface

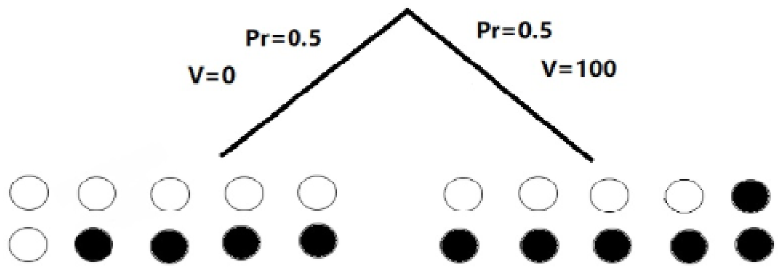
There was a three-page instruction placed on the table for each subject before they enter the lab. After all the subjects were seated, they had five minutes to read the instruction. The instruction includes the basic idea of the game, explanation of the mechanism of the elicitation method, and how the payoffs are determined. An experimenter was answering any questions subjects may have after the reading. Then the experimenter did a demonstration with two identical bags and real bingo balls that are colored either dark blue and white to walk through the whole process again to make sure that all subjects understand the game thoroughly. After the demonstration, there was a ten minutes presentation with power points to familiarize subjects with the computer user interface. At the end of the presentation, there were some quiz questions (see appendix A.4.3). Based on the subjects' answers, it was clear that all of the subjects understood the game and mechanism well.

Figure 1.2 shows the user interface for the first draw (Other user interfaces are provided in section A.1.1). The figures for the baseline and second draws are similar to figure 1.2. On the top, subjects can observe the timer; they all have twenty seconds for each decision page. Below the timer, subjects can observe the current round number. The figure in the middle reminds subjects of the prior and how the balls are placed in the two urns. After the balls are drawn from the urns, the computer provides both the number of black (also white) balls and a visual display of balls in a format similar to the prior. Before clicking the "Next" button, the User Interface reminds the subjects that their best strategy is to report truthfully.

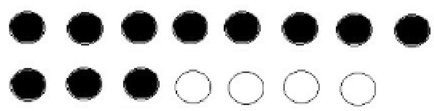
First Draw

Time left to complete this page: **0:12**

Current round number: 1



You still don't know the true value of the lottery. However, the computer has drawn 15 balls with replacement in the bag that corresponds to the determined lottery value. Within those 15 balls that just drawn, they contain 11 black balls and 4 white balls.



Again, please enter the "chances out of 100" that the true value of the lottery is 100.

Your stated probability:

Note: It is in your BEST interest to state your TRUE value.

Next

Figure 1.2: User Interface for the First Draw

After each round, the payment for this round is provided in the "Round Payment" page. Subjects have fifteen seconds to read this page before the next round begins. After clicking the "Next" button for the next round, there will be a waiting page so that all players go to the next page at the same time. After all fifty rounds were finished, subjects will be directed to the "Result" page for their final payment of this session.

1.4.3 Determinant of Payment

In addition to the $7 U.S. show-up fee, each decision subject makes in the experiment will result in the possibility of earning money. For the three decisions subjects made in each round, the computer will roll a single six-sided die. If one dot comes, decision 1 will be the one that counts for their payment. If two dots or three dots come, then decision 2 will be the one that counts. If four or five or six dots come, then decision 3 will be the one that counts. After they completed all 50 rounds for the game, the computer will randomly draw a number between 1 and 50 to decide which round will be chosen to be the one for their final payment in this game. The real earning a subject receives will be her final payoff in the game multiplied by 0.15.[4] Subjects' payoff is from either the initial lottery or the N-lottery, therefore, their payoff is either 0 or 100; after multiplied by the convert rate, 0.15, they could earn either $0 or $15 from the game, plus their $7 show-up fee, subjects can earn either $7 or $22 for the whole game.

[4] 0.15 is the convert rate. It is calculated to make sure each subject can make a reasonable payoff in a 45 to 50 mins session.

1.5 Results

1.5.1 Comparison of Models

This subsection is the main result of this paper and it plans to do a comparison of different models to find out which model fits the data better. The two most important models to be compared are the Bayesian Model and the Average model, which are introduced in the previous sections. Here is a quick look of the other three models.

First is the “Linear in the Ratio Model”. The first (resp. second) signal draws n_1 (resp. n_2) balls of the corresponding bag and k_1 (resp. k_2) of them are black balls and m_1 (resp. m_2) of them are white balls. Instead of assuming people process each signal in a perfect Bayesian way (this is the assumption of the Average Model), the “Linear in the Ratio Model” assumes people simply calculate the ratio of the black balls in a draw, that is calculating k_1/n_1 and k_2/n_2. Therefore, the “Linear in the Ratio Model” assumes people's posterior after observing a single signal is a function of the ratio: posterior $= f(k_1/n_1)$.

The second is the “Weighted Probability Function”. This model also relaxes the assumption that people process a single signal in a Bayesian way, it borrows the weighting function from Tversky and Kahneman (1992) to reweight the Bayeisan posterior after each signal:

$$w_i(p) = \frac{p^{\gamma_i}}{(p^{\gamma_i} + (1-p)^{\gamma_i})^{1/\gamma_i}}, \text{ where } w_i(p) = p \text{ when } \gamma_i = 1$$

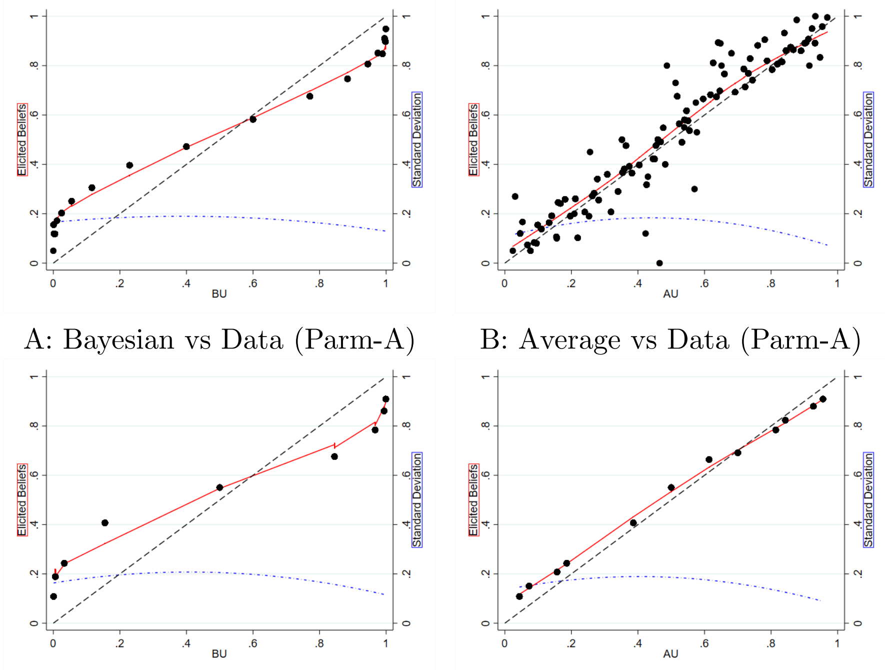
Where, p is the Bayesian posterior given one of the signals, $w_i(p)$ is the re-weighting function for individual i and γ_i is the parameter to be estimated given the data.

The third model is the "Bayesian Multiplication" model. This model maintains the assumption that people process each single signal in a perfect Bayesian way and thus obtain the posterior of p_1 and p_2 given each signal. However, the difference is that this model assumes people take the multiplication of the two posteriors to obtain their final posterior given two signals: $\text{Bayes\_multiplicate} = p_1 \times p_2$.

Section 1.5.1.1 discusses the results of the Bayesian Model and the Average Model. Section 1.5.1.2 discusses the Linear in the Ratio model and section 1.5.1.3 talks about "Weighted Probability Function Model" and "Bayesian Multiplication Model".

1.5.1.1 Bayesian Model and Average Model

Figure 1.3 helps us to visualize which models' (Bayesian or Average) predicted probabilities fit the elicited data better. The x-axis is the model prediction if we plug the number of black balls and white balls in both signal 1 and signal 2 into the Bayesian (denoted as "BU" on the left two graphs) or Average models (denoted as "AU" on the right two graphs). The y-axis on the left side is the elicited probabilities after subjects observing both signals. Given a specific pair of signals, models have one unique prediction, but subjects may have different posterior probabilities; I calculated the average and the standard deviation of those elicited posteriors. The black dots on the graphs are the averaged elicited beliefs. The standard deviations (also the right y-axis) for these elicited beliefs are represented by the fitted blue dash-dotted line, which is a quadratic fit. The black dashed line is the 45-degree line and the red curve is the lowess fit for the elicited posteriors. The top two graphs are for Parm-A sessions, while the bottom two are for Parm-B sessions.



A: Bayesian vs Data (Parm-A) B: Average vs Data (Parm-A)

C: Bayesian vs Data (Parm-B) D: Average vs Data (Parm-B)

Note: The left y-axis (data) is subjects' posterior beliefs after observing both signals. Each dot is the average over all subjects and all sessions. The black dashed line is the 45-degree line and the red line is the lowess fit. Figure A4 plots the second order polynomial fit with 95% confidence interval and it gives us similar information.

Figure 1.3: Goodness of Fit: Bayesian vs Average

If the model perfectly predicts subjects' elicited posteriors, the dots should lie on the dashed 45-degree line. Graph A and C in figure 1.3 shows an inverse-S-shaped pattern as in Tversky and Kahneman (1992) and Holt and Smith (2009). Bayes rule under-predicts subjects' posterior when the probability is lower than 0.5 and over-predicts when probability is higher than 0.5. For all observations, I ran the regression of (Elicited Probabilities) $= \beta_0 + \alpha_{ID} + \beta_1$(Bayes Prediction) with individual fixed effect and clustering in session level. Regression results are on column 1 (Parm-A) and column

5 (Parm-B) of Table 1.1. For both regressions on Bayesian model, the intercepts are higher than 0, and the slopes of the fitted lines are flatter than the 45-degree line. In the regression, the constant term is 0.126 and 0.338 respectively, and the slopes are 0.647 and 0.601. At the bottom of the table, all p-values for the hypothesis (1) $H_0 : \beta_0 = 0$; (2) $H_0 : \beta_1 = 1$ and (3) the joint test ($H_0 : \beta_0 = 0, \beta_1 = 1$) are low for both regressions, meaning that Bayes' prediction significantly deviates from 45-degree line when predicting people's actual behavior.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Sessions	Parm-A	Parm-A	Parm-A	Parm-A	Parm-B	Parm-B	Parm-B	Parm-B
VARIABLES	Data	Data	Data	Data	Data	Data	Data	Data
Bayesian	0.647***				0.601***			
	(0.0525)				(0.0386)			
Average		0.991***				0.870***		
		(0.0734)				(0.0558)		
LR(Ave)			1.370***				0.864***	
			(0.0959)				(0.0573)	
LR(Bayes)				0.820***				0.574***
				(0.0602)				(0.0429)
Constant	0.126**	-0.0443	-0.255**	0.0177	0.338***	0.202**	0.189**	0.337***
	(0.0259)	(0.0363)	(0.0490)	(0.0314)	(0.0220)	(0.0307)	(0.0327)	(0.0258)
Observations	1,360	1,360	1,360	1,360	1,720	1,720	1,720	1,720
R-squared	0.738	0.747	0.741	0.748	0.709	0.734	0.730	0.667
RMSD	0.1506	0.1482	0.1500	0.1478	0.1603	0.1532	0.1544	0.1714
p for $\beta_0 = 0$	0.0082	0.2892	0.0065	0.6041	0.0001	0.0028	0.0044	0.0002
p for $\beta_1 = 1$	0.0025	0.9118	0.0182	0.0401	0.0005	0.0806	0.0762	0.0006
p for joint test	0.0025	0.9118	0.0182	0.0401	0.0005	0.0806	0.0762	0.0006

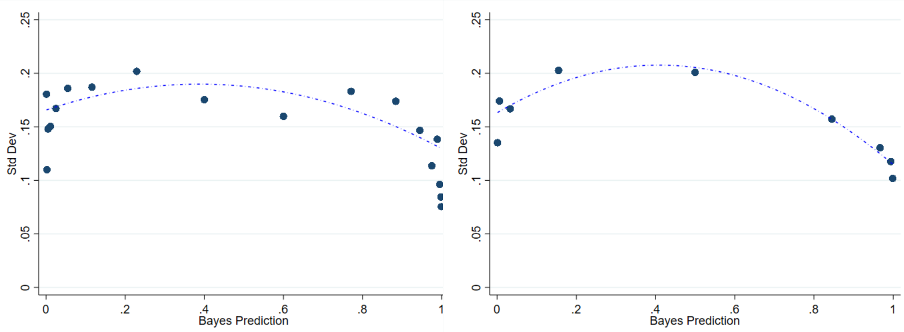
Robust standard errors in parentheses
*** p<0.001, ** p<0.01, * p<0.05

Note: Observations are calculated as the following: the first and the last 5 rounds were deleted in case of heuristics at the begining and the end; and it results 40 rounds of data. For Parm-A sessions, 34 subjects with 40 rounds gives us 34*40=1360 observations. For Parm-B sessions, we have 43*40=1720 observations. Also, please note that the p-value for $\beta_1 = 1$ is the same as the p-value for the joint test of $\beta_0 = 0$ & $\beta_1 = 1$. The consulting results with one of the econometrician (A special thanks to Professor Jessie Li's help.) is that in the joint test, $\beta_0 = 0$ part will eventually cancel out and the test would be the same as if conducting test for $\beta_1 = 1$. The STATA result also shows "constraint 1 dropped" when performing the joint test.

Table 1.1: Regressions for Goodness of Fit

Graph B and D in figure 1.3 shows the Average model's prediction against elicited probabilities. From these graphs, there is no obvious pattern like that in graph A and C. The points lie both above and below the 45-degree line. The lowess fit looks like a linear fit plus noise. I ran the same regression for Average predictions and the results are on column 2 (Parm-A) and column 6 (Parm-B) of Table 1.1. The regression lines are, relative to Bayesian's prediction, closer to the 45-degree line with a constant term close to 0 and slope close to 1. The hypothesis tests for $\beta_0 = 0$, $\beta_1 = 1$ and the joint test tell us that Average model does especially well in Parm-A sessions, the test results indicate that regression of elicited data on Average models is a regression through origin with slope of 1. Average model's prediction does not perform as well in Parm-B sessions as in Parm-A sessions, the regression result from column 6, including the hypothesis tests, shows us that the fitted line is flatter than the 45-degree line.

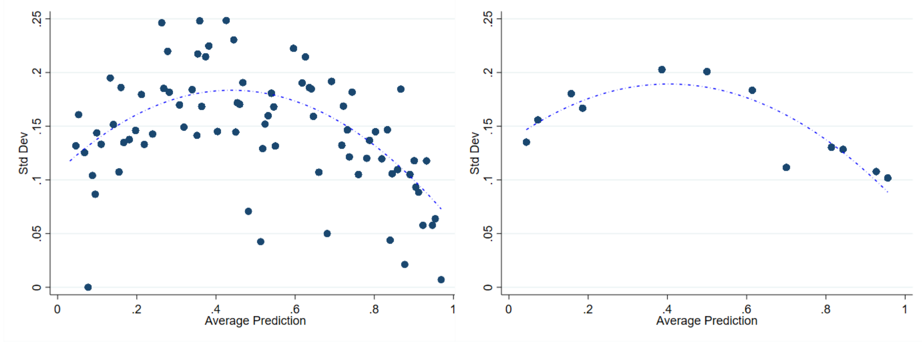
A potential reason for the flatter fit of Average model in parm-B sessions is that there are some observations located at a lower probability when Average model's prediction is high, and there are some located at a higher probability when Average model's prediction is low. These "outliers" are dragging the fitted line down to a flatter pattern. But why there are some "outlier" observations in parm-B sessions? Although the calculation of Shannon's Entropy tells us that the parm-A sessions' uncertainty is similar to that of parm-B sessions', 15 and 10 protocol provides subjects more information. In parm-B sessions, with less balls drawn from the urn, subjects might feel less informed by the signals and thus their elicited belief tend to be more conservative given a specific pair of signals. In another words, when the model's prediction is high (low), the subjects might tend to report a lower (higher) probability since they feel less confident relative to parm-A sessions. If this is the case, we should be able to observe higher standard deviations at the two end points in parm-B sessions. In figure 1.3, the quadratic fits are in an inversed-U shape, this shape makes sense because when the predicted posteriors are at 0.5, people should be more uncertain about the probability and thus the standard deviation of elicited beliefs should reach its maximum. Figure 1.4 and 1.5 below included the standard deviations as dots so that we could get a more clear pattern. The left graphs are from the parm-A sessions, they show us that given each predicted posterior from Bayes and Average models, what is the standard deviation of the elicited posteriors correspond to that pair of signals, the blue dash-dotted curve is the quadratic fit. Similarly, the right graphs are from the parm-B sessions. It is clear that the blue line is more curved in the parm-A sessions. This means, when we are at the two ending points (0 and 1), the standard deviation in the parm-A sessions is smaller than the standard deviation in the parm-B sessions. Such difference indicates elicited posteriors at the two ending points are noisier in the parm-B sessions. This higher standard deviation provides an explanation for the Average models' poorer performance at parm-B sessions.



A: Parm-A B: Parm-B

Figure 1.4: Standard Deviation for Elicited Data Given Each Bayes Prediction

So, according to the above regression analysis, Average model outperforms



A: Parm-A B: Parm-B

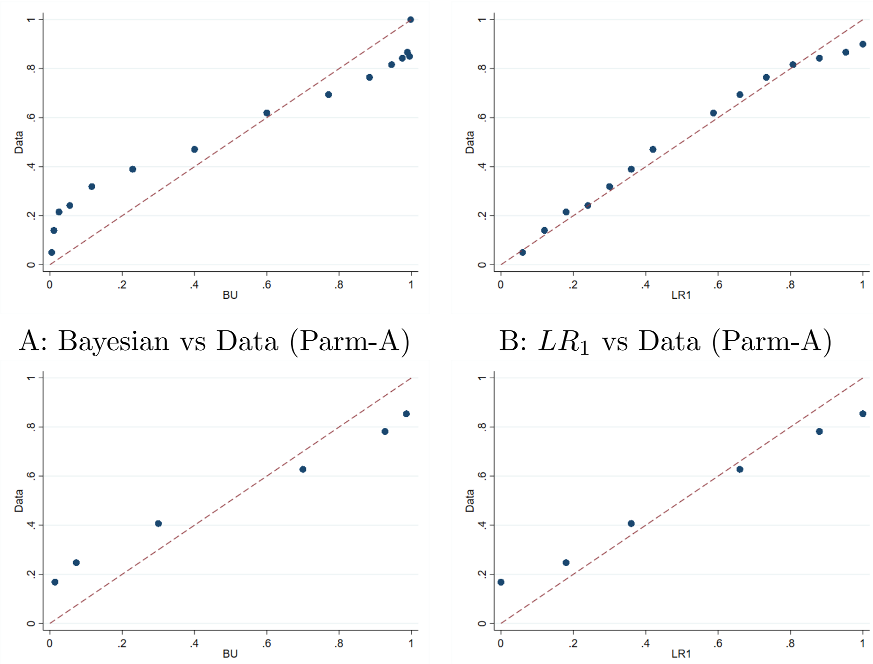
Figure 1.5: Standard Deviation for Elicited Data Given Each Average Prediction

Bayesian model in both Parm-A and Parm-B sessions. Although Average model outperforms the Bayesian model, we might have subjects assign different weights on signals they observe. That means other than $Average = 0.5p_1 + 0.5p_2$, we might have $Average = wp_1 + (1 - w)p_2$. Whether there exist other pairs of weights that also fit the data well remains unknown. To find out, I changed the weight on the first draw from 0.1 to 0.11, 0.12, ... , 0.89, 0.99, and the results show us that for Parm-A sessions, there is a pretty wide range of weights (0.18,0.75) that fit the data well. For those pair of weights, the intercept is close to 0, slope is close to 1. This indicates recency bias might play a role here. For Parm-B sessions, however, the weight range that fit the data well narrowed down to (0.38,0.5). The reason for the above difference in the two sessions could resulted from the reduction of balls drawn from the bag in Parm-B sessions. With less balls drawn, given a specific pair of signals, the elicited probability might not disperse as much as in Parm-A sessions. Although the standard deviation is higher for Parm-B sessions in the two ending points 0 and 1, figure 1.5 shows us that the points of standard deviation in Parm-A sessions are more spread out in general. Such disperse in Parm-A sessions creates a wider range of weights that might fit the data well.

1.5.1.2 Linear In The Ratio Model

As we could see from the Average model, it has a strong assumption and that is it assumes people process one single signal in a perfect Bayesian way. Many works like Tversky and Kahneman (1974), Tversky and Kahneman (1971), and Holt and Smith (2009) had shown it is not the case. Graph A and C in figure 1.6 shows us Bayesian prediction vs elicited data for signal one only. The two graphs on the left side exhibit the same inverse-S-shaped pattern as in the previous works, which directly points out that the assumption of Average model is too strong. To relax this assumption, I will introduce a model derived from subjects' feedback and I call it "Linear in the Ratio" (LR) model.

In the post-experiment survey, subjects need to write down how do they report their beliefs. Many subjects use a more straightforward way of processing the single signals and calculate their posterior beliefs. What they do is calculating the ratio of the black balls in a draw. For example, in Parm-A sessions, if signal one provides 10 black balls, subjects will calulate the proportion of the black balls, which is 10/15. According to some of the subjects' numerical examples, their posterior actually increases faster than the ratio, so their posterior belief is a little higher than 10/15. So, I think when the ratio of the black balls is lower than 0.5, their posterior would decrease faster than



A: Bayesian vs Data (Parm-A) B: LR_1 vs Data (Parm-A)

C: Bayesian vs Data (Parm-B) D: LR_1 vs Data (Parm-B)

Note: The Y-axis (data) is subjects' posterior beliefs after the first signal. Each dot is averaged over all subjects and all sessions.

Figure 1.6: Two Updating Rules vs Data (Signal 1 Only)

the ratio as well. Therefore, I have the following "Linear in the Ratio" (LR) model.

$$LR_1 = \begin{cases} 1.1 * (\text{blk}/15), \text{ if blk} >= 8 \\ \\ 0.9 * (\text{blk}/15), \text{ if blk} <= 7 \end{cases} \qquad LR_2 = \begin{cases} 1.1 * (\text{blk}/10), \text{ if blk} > 5 \\ \\ 0.5, \text{if blk} = 5 \\ \\ 0.9 * (\text{blk}/10), \text{ if blk} < 5 \end{cases} \tag{1.4}$$

Where "blk" means "number of black balls", LR_1 and $LR_2 = 1$ if LR_1 and $LR_2 > 1$. LR_i means subjects' posterior after observing signal i. The above equations 1.4 are for Parm-A sessions, that's why we have number of black balls divided by 15 and 10, since those are the number of balls drawn from the bag. Parameters 8, 7, and 5 are used to separate the ratio that is higher than 0.5 from that is lower than 0.5. Similarly, equations 1.5 are the equations for Parm-B sessions.

$$LR_1 = \begin{cases} 1.1 * (\text{blk}/5), & \text{if blk} >= 3 \\ \\ 0.9 * (\text{blk}/5), & \text{if blk} < 3 \end{cases} \qquad LR_2 = \begin{cases} 1.1 * (\text{blk}/3), & \text{if blk} >= 2 \\ \\ 0.9 * (\text{blk}/3), & \text{if blk} < 2 \end{cases} \tag{1.5}$$

Where, again, “blk” means “number of black balls”, LR_1 and $LR_2 = 1$ if LR_1 and $LR_2 > 1$. Please note that this experiment did not have the data of subjects’ posterior for only observing signal two. Therefore, to test the LR model, what I use is the observations for signal one only. Graphs B and D in figure 1.6 shows us how this LR model fits the data for signal one. As we could see from graph B, which is for Parm-A sessions, other than the model is a bit over predicting the data when the probability is high, the rest parts fit pretty well. For graph D of Parm-B, however, the model underpredicts when probability is low, but overpredicts when probability is high, the reason could be that the number of balls drawn from the bag reduces a lot in Parm-B sessions, thus subjects tend to be less confident of the posterior probability, and thus they reported relatively more conservative beliefs. The alternative reason could be the parameter for LR model in parm-B sessions should be different. Note I have parameters 1.1 and 0.9 in the LR model to fit subjects’ feedback, these parameters might be different in parm-B sessions. But in order to have the model being consistent, I choose to keep these two parameters the same for both parm-A and parm-B sessions.

From the above analysis for signal 1, we get an idea of how the LR model fit the data for signal 1. The next step is to use the LR model to estimate subjects’ posterior for observing both signals and test the goodness of fit. LR_2 is the estimated probability assuming subjects only observe signal 2. I need to combine LR_1 and LR_2 so that I could have LR model’s prediction of posteriors for both signal 1 and signal 2.

There are two ways of combination I tested: (1) combine in the Average way, and (2) combine in the Bayesian way. Average way of combining is straightforward:

$$\text{LR(Ave)} = 0.5 * LR_1 + 0.5 * LR_2 \tag{1.6}$$

Bayesian way of combining is based on Massaro and Friedman (1990). According to Massaro and Friedman (1990), LR_1 and LR_2 can be think as the values supporting one of the alternatives (lottery worth 100). The goodness of match with a LR_1 and LR_2 alternative can be represented by:

$$a_{lotteryworth100} = LR_1 \times LR_2 \quad \& \quad a_{lotteryworth0} = (1 - LR_1) \times (1 - LR_2)$$

Given the above two values, the prediction of using Bayesian combination is:

$$\text{LR(Bayes)} = \frac{a_{lotteryworth100}}{a_{lotteryworth100} + a_{lotteryworth0}} \tag{1.7}$$

Massaro and Friedman (1990) calls the Bayesian way of combination Fuzzy-Logical Model of Perception (FLMP), and they pointed out that the Bayes' theorem and the FLMP are conceptually equivalent if we have the assumption that the fuzzy-truth value conincide that of subjective conditonal probability.

Now, we have another two predictions (LR (Ave) and LR (Bayes)). I ran the regression like the previous models and the results are in column (3), (4), (7), and (8) in table 1.1. The results show that combine LR_1 and LR_2 in Average way gives us a steeper fitted line relative to combine in Bayes way. P-values for the hypothesis testing tell us that LR (Ave) does a relatively better job fitting the data than LR (Bayes). LR (Ave) model performs very well for Parm-B sessions if we drop the outliers (see table A1). RMSD in table 1.1 shows us that none of the models tested so far outperforms other models significantly.

1.5.1.3 Weighted Probability Function and Bayesian Multiplication

Another two models that this subsection aims to test are Bayesian Multiplication models and Weighted Probability Function. Bayesian Multiplication model is straightforward. Average model says subjects would take the average of p_1 and p_2, Bayesian Multiplication model says they might take the multiplication:

$$\text{Bayes\_multiplicate} = p_1 \times p_2$$

In Tversky and Kahneman (1992), they re-weighted the Bayesian probability so that the re-weighted probability exhibits inverse-S-shaped pattern. The weighting function is in below.

$$w_i(p) = \frac{p^{\gamma_i}}{(p^{\gamma_i} + (1-p)^{\gamma_i})^{1/\gamma_i}}, \text{ where } w_i(p) = p \text{ when } \gamma_i = 1 \tag{1.8}$$

Function 1.8 has the assumption of heterogeneity, which assumes each subject might have different γ, denote as γ_i. In order to find the best fitted γ_i, I used the data from signal 1, since subjects only observe signal 1 and were not affected by signal 2, in which case could help us to find γ_i for each subjects. For each subject, I select $\gamma_i = 0.1, 0.11, 0.12, ..., 0.99$ and calculated the sum of squared error ($SSE_i = \sum(p_{elicited} -$

$w_i(p))^2$). Then I found the best estimate of γ_i for each individual, which is the γ_i that minimizes SSE_i. The majority of the best γ_i in both Parm-A and Parm-B sessions fall in the range of (0.6,0.8). I also did weighted probability assuming homogeneity, which means I assume all subjects have the same γ parameter, SSE tells us the best $\gamma = 0.72$ for Parm-A, and $\gamma = 0.65$ for Parm-B sessions.

Now, we have Weighted Probability for signal 1 (denote as $WeighProb_1$). To test this Weighted Probability Function, I assume subjects process the second signal the same way as the first signal, which I denote their posterior probability of signal 2 as $WeighProb_2$. Then, similar as what have been done for LR model, I combine $WeighProb_1$ and $WeighProb_2$ in both Average way and Bayesian way to get two different predictions. Please note that I have Weighted Probability assuming both heterogeneity and homogeneity. Therefore, what I am going to test are four predictions: (1) WeighProb_hete(Ave), (2) WeighProb_hete(Bayes), (3) WeighProb_hom(Ave), and (4) WeighProb_hom(Bayes). Together with Bayes_multiplicate, the following regression table 1.2 shows us the results for the five predictions' goodness of fit.

From table 1.2, we could see that for weighted probability function predictions, the fitted line is steeper if it is combined in Average way, and the fitted line is flatter if it is combined in Bayes way. The reason the fitted line is flatter for Bayes combination is relatively more straightforward. As mentioned in Massaro and Friedman (1990), Bayes's theorem could be considered conceptually equivalent to the FLMP model, therefore, when we combine weighted probability function in the Bayes way and compare the prediction with elicited data, we should be able to see a similar inverse-S-shaped pattern as if we use perfect Bayesian model, and thus the fitted line is flatter. The reason that the fitted line is steeper for Average combination is not clear. Although we re-weight Bayes prediction (p_1 and p_2) to fit individuals' elicited observations for each signal, but such re-weighting does not capture the exact process when subjects were processing the signals. Therefore, the best fitted γ_i should be doing a good job in a relative sense, not in an absolute sense. Bayes multiplication model also provides a flatter fit. Although the p-value for joint test of Bayes multiplication model tells us that it does a relatively good job fitting parm-A sessions' data, RMSD illustrates that none of the models did a significant better job than others. In the next section, I will talk about analysis for individual behaviors.

1.5.2 Individual Behavior

Next question of interest is how individual behaves. Are they all consistent with Bayesian, or with Average, or neither? To begin with, one thing we could look at are the four cases p_1 and p_2 could be combined:[5]

Case 1: Both p_1 and p_2 are greater than 0.5.

Case 2: p_1 is greater than 0.5, but p_2 is smaller than 0.5.

Case 3: p_1 is smaller than 0.5, but p_2 is greater than 0.5.

Case 4: Both p_1 and p_2 are smaller than 0.5.

The cases that could help us distinguish who is Bayesian and who is Average are case 1 and case 4 and thus these two cases are the first focus for this subsection, later I will provide results for all cases combined. Based on the Bayes equation in section 3,

[5]As a reminder, $p_1 = BU(V = 100|k_1, m_1)$, and $p_2 = BU(V = 100|k_2, m_2)$, where k_1 (m_1) means number of black (white) balls in the first signal, k_2 (m_2) means the number of black (white) balls in the second signal. BU means Bayes equation.

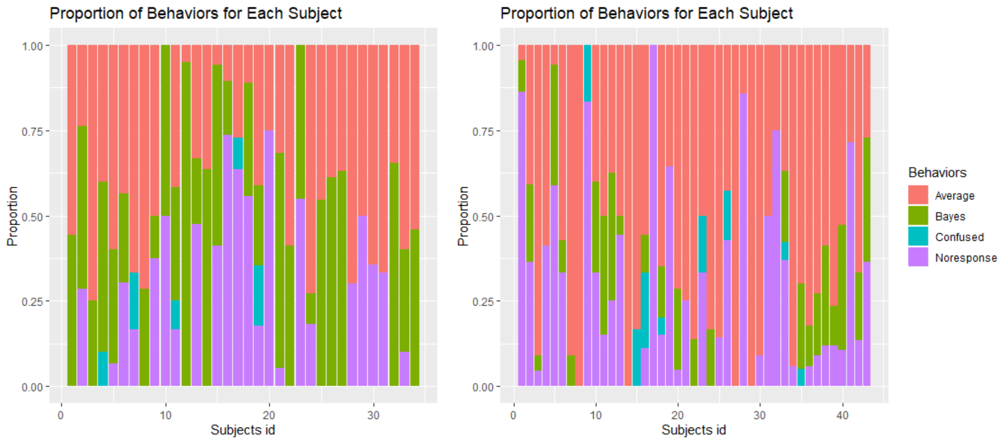
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Sessions	Parm-A	Parm-A	Parm-A	Parm-A	Parm-A	Parm-B	Parm-B	Parm-B	Parm-B	Parm-B
VARIABLES	Data	Data	Data	Data	Data	Data	Data	Data	Data	Data
WeighProb_hete(Ave)	1.277***					1.144***				
	(0.0658)					(0.0325)				
WeighProb_hete(Bayes)		0.790***					0.730***			
		(0.0490)					(0.0250)			
WeighProb_hom(Ave)			1.246***					1.153***		
			(0.0910)					(0.0719)		
WeighProb_hom(Bayes)				0.761***					0.724***	
				(0.0592)					(0.0453)	
Bayes_multiplicate					0.887***					0.728***
					(0.0718)					(0.0430)
Constant	-0.128*	0.115**	-0.136*	0.105*	0.204***	0.0573*	0.269***	0.112*	0.322***	0.426***
	(0.0296)	(0.0205)	(0.0424)	(0.0265)	(0.0195)	(0.0177)	(0.0142)	(0.0356)	(0.0224)	(0.0150)
Observations	1,360	1,360	1,360	1,360	1,360	1,720	1,720	1,720	1,720	1,720
R-squared	0.771	0.771	0.741	0.741	0.670	0.743	0.732	0.727	0.720	0.637
RMSD	0.1408	0.1409	0.1499	0.1500	0.16907	0.1506	0.1538	0.1551	0.1573	0.1789
p for $\beta_0 = 0$	0.0122	0.0050	0.0329	0.0168	0.0005	0.0318	0.0000	0.0349	0.0001	0.0000
p for $\beta_1 = 1$	0.0135	0.0129	0.0538	0.0156	0.1914	0.0114	0.0004	0.1007	0.0037	0.0032
p for joint test	0.0135	0.0129	0.0538	0.0156	0.1914	0.0114	0.0004	0.1007	0.0037	0.0032

Robust standard errors in parentheses
*** p<0.001, ** p<0.01, * p<0.05

Table 1.2: Regressions for Goodness of Fit of Weighted Probability Function and Bayesian Multiplication

it is easy to show that, if it is case 1, Bayesian posterior should be strictly higher than both p_1 and p_2. However, in the Average model, the posterior will lie between p_1 and p_2 since it is the average of both. For example, if the first signal provides 10 black balls and 5 white balls, while the second signal provides 6 black balls and 4 white balls. We'll have $p_1 = 88.4\%$ and $p_2 = 69.2\%$. This will gives us Bayesian posterior of 94.5%, which is higher than both p_1 and p_2. While the Average posterior of 78.8%, which lies between p_1 and p_2. Similarly, if it is in case 4, Bayesian posterior should be strictly smaller than both p_1 and p_2 and Average posterior should lie between them. Therefore, case 1 and case 4 could help us capture whether subjects suffer from representative bias. We do not look at case 2 and case 3 here is because in those two cases, Bayesian update and Average update might both lie between p_1 and p_2, thus it is hard to distinguish whether subjects are updating based on Bayesian or Average. In addition to Average behavior and Bayesian behavior, I also categorized another two possible behaviors: No-response and Confused. No-response defines the behavior when, after observing both signals, subjects' third elicited belief stays the same as the second one even if p_1 and p_2 differs (in another words, No-response means subjects do not respond to the second signal). Confused defines when subjects' third elicited belief goes to the opposite direction while p_2 supports what p_1 suggested (in another words, Confused means subjects react to the second signal to the wrong direction).

Figure 1.7 shows us the proportion of the four behaviors among individuals for both Parm-A and Parm-B sessions. X-axis is subjects' id, Y-axis is the proportion of the four behaviors, where the longer the bar, the more that individual exhibit the corresponds behavior in case 1 and case 4. Among all 34 subjects in Parm-A sessions, two



A: Parm-A Sessions B: Parm-B Sessions

Figure 1.7: Proportion of Behaviors for Individuals (Case 1 and Case4)

subjects are consistent with Bayesian updating rule and never exhibit Average behavior. Seven subjects are consistent with the Average updating rule and never exhibit Bayesian behavior. But for the rest of the subjects, everyone shows some Bayesian updating and also Average updating. Also, although we don't see a substantial proportion of Confused behavior, but we do have many subjects that exhibit No-response behavior. For Parm-B sessions, the results are little different. While the proportion of Confused behavior stays similar to that in Parm-A sessions, the proportion of No-response behavior increased in Parm-B sessions. Also, Bayesian behavior decreased because the Average behavior increase. The reason of such change in Parm-B sessions could be that the LR behavior increased. In Parm-B sessions, if I compare Average model's predictions with LR model's predictions, I found out that, Average model's predictions are very similar to LR model's predictions. According to subjects' feedback, in parm-B sessions, the reduce of balls drawn from the bag also reduced the difficulty for the subjects' to calculate the proportion of the black balls in a signal, therefore, LR behavior increased more in Parm-B sessions. Thus the LR behavior increased and it explains why Average behavior increased in graph B of figure 1.7 since these two models have similar goodness of fit in Parm-B sessions.

Therefore, at an individual level, we can not conclude that people are consistent with one of the behaviors, their updating process is more mixed.

1.5.3 Learning Patterns and Expected Payoffs

This subsection tries to find out whether subjects exhibit any learning patterns. The goal is to find whether, as the rounds go by, subejects behave more like Bayesian or Average. To begin with, I use data of case 1 and case 4 as in the previous section, and define behaviors as Bayesian or Average depending on the comparison of subjects' elicited belief on p_1 and p_2.

Table 1.3 is the regression of estimating the effect of rounds on subjects' behavior (Bayesian or Average). The regression includes only case 1 and case 4, therefore we have 707 observations for Parm-A sessions and 1204 observations for Parm-B sessions.

$$Bayesian/Average = \beta_0 + \alpha_{ID} + \beta_1 Round$$

The outcome variable "Bayesian" and "Average" are both binary variables, indicating that if subjects behave like Bayesian, we'll have $Bayesian = 1$, otherwise 0. Also, if subjects behave like Average, we'll have $Average = 1$, otherwise 0. The regression estimates the effect of rounds while controlling the individual fixed effect, and clustering at session level. The result in regression 1 shows that our explanatory

Sessions VARIABLES	(1) Parm-A Bayesian	(2) Parm-A Average	(3) Parm-B Bayesian	(4) Parm-B Average
Round	-0.000204	-0.00436*	0.00108	0.000125
	(0.000539)	(0.00154)	(0.000846)	(0.00107)
Constant	0.369***	0.572***	0.0797*	0.0325
	(0.0146)	(0.0417)	(0.0216)	(0.0274)
Observations	707	707	1,204	1,204
R-squared	0.249	0.130	0.102	0.189

Robust standard errors in parentheses
*** $p<0.001$, ** $p<0.01$, * $p<0.05$

Table 1.3: Round Effect on Behaviors (Case 1 and Case 4)

variable "Round" does not have a statistically significant effect on being Bayesian for Parm-A sessions. However, regression 2 shows that as each additional round finished, subjects are 0.436 percentage point less likely to behave like Average. For Parm-B sessions, however, the results show there is no converge to either of the behaviors over time.

For Parm-A sessions, subjects become less likely to behave like Average should not be resulted from learning Bayesian rule as a more accurate prediction of the true status. Because the sessions do not provide subjects the true value of the lottery at the end of each round. However, the sessions do provide the payoff for each of their decisions at the end of each round. Therefore, the reason that subjects are less likely to be Average might be resulted from learning that being Bayesian will result in a higher expected payoff in this game. I proved why being a Bayesian will have a higher expected payoff, and the proof is provided in the appendix section A.3.2.

The above analysis is done by using case 1 and case 4 data, the following analysis shows us the pattern, if any, if we include all four cases. To find out whether subjects are becoming closer to Bayesian or Average, I calculated the absolute differences between subjects' posterior belief and predictions from two models after observing two signals:

$$|\text{Elicited Belief} - \text{Bayesian's Prediction}| \text{ and } |\text{Elicited Belief} - \text{Average's Prediction}|$$

I borrow the regression method from Noussair et al. (1995), which is often used to study convergence in experimental data, to study the convergence trend for my data. The model assumes the dependent variable exhibits a convergence pattern over time within each experimental session and approaches some value asymptotically. Here, the dependent variables are the above two differences, and the model can help us to find out if one of them is approaching to zero overtime.

$$\begin{aligned} y_{it} =& B_{11}D_1(1/t) + B_{12}D_2(1/t) + B_{13}D_3(1/t) + B_{14}D_4(1/t) + B_{15}D_5(1/t) + \alpha_{ID} \\ &+ B_2(t-1)/t + u \end{aligned}$$

Where i indicates the particular experiment session (5 sessions in my experiment). t represents rounds that each individual report their beliefs. D_i is a dummy variable that takes a value of 1 for session i and a value of 0 otherwise. B_{1i} indicates the origin of the dependent variable. Please note that if $t = 1$ then the value of the dependent variable equals to B_{1i}, but when t increases, the weight for B_{1i} reduces because of $(1/t)$, and the weight of B_2 gets larger because $(t-1)/t$ approaches 1. Therefore, B_2 is the asymptote of the dependent variable and thus it is our main focus for the regression result.[6]

Dependent Var	B_{11}	B_{12}	B_{13}	B_{14}	B_{15}	B_2	Hypothesis	Significance(p)	R^2
Parm-A									
$\|Dta - Bayes\|$	0.06*	0.12**	0.12*	0.05	0.07	0.11***	$H_0 : B_2 = 0$	< 0.001	0.6627
	(0.03)	(0.04)	(0.06)	(0.06)	(0.04)	(0.02)			
$\|Dta - Ave\|$	0.05*	0.06	0.04	0.04	0.05*	0.08***	$H_0 : B_2 = 0$	< 0.001	0.7087
	(0.02)	(0.03)	(0.05)	(0.04)	(0.02)	(0.01)			
Parm-B									
$\|Dta - Bayes\|$	0.24***	0.22***	0.18**	0.20***	0.25***	0.18***	$H_0 : B_2 = 0$	< 0.001	0.691
	(0.06)	(0.06)	(0.05)	(0.05)	(0.06)	(0.04)			
$\|Dta - Ave\|$	0.29***	0.23***	0.25***	0.25***	0.27***	0.25***	$H_0 : B_2 = 0$	< 0.001	0.6296
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)			

Robust standard errors in parentheses
*** p<0.001, ** p<0.01, * p<0.05

Table 1.4: Regressions for Beliefs Convergence

In the regression table, I denote |Elicited Belief − Bayesian's Prediction| as $|Dta - Bayes|$, and denote |Elicited Belief − Average's Prediction| as $|Dta - Ave|$. Testing of the convergence is focused only on the variable B_2, which represents the long-term (asymptotic) tendency of the magnitude of the dependent variables. The result from table 1.4 shows that, for both Parm-A and Parm-B sessions, we can reject hypothesis $H_0 : B_2 = 0$ for both of the regressions. The coefficient B_2 is significant for both dependent variables with a p-value smaller than 0.01, which indicates that subjects' posterior belief converges to neither of the model's prediction in the long-term.

1.6 Conclusion

Inspired by Anderson (1981), Tversky and Kahneman (1974), Epstein et al. (2010), and Holt and Smith (2009), this paper introduced an Average model and de-

[6]In Noussair et al. (1995)'s regression, they corrected for first-order autocorrelation, which is necessay for their experiment because they study the pattern of the market price. However, I don't think this is a concern for this experiment, because each round's true lottery value and the corresponding balls draw from the urn are independent from each other. So in this regression, I corrected for heteroscedasticity only.

signed a lab experiment to test the performance of the Average model in terms of predicting subjects' posterior belief when they are facing an uncertain event accompanied by two signals. The hypothesis is that the Average model will outperform the Bayesian rule because (1) it can capture representative bias (Tversky and Kahneman (1974)), and (2) this model requires less complication to process the information. Using Karni (2009)'s theory, subjects' posterior beliefs were elicited in terms of "chances out of 100" that their lottery worth 100.

The comparison between Average model and Bayesian model tells us that the Average model outperforms Bayesian model in terms of the goodness of fit. In the regression of elicited probability on Bayesian predictions, the graph shows an inverse-S-shaped form indicating Bayesian under predicts when the probability is low and over predicts when the probability is high. However, when comparing data with the Average model's predictions, there is no obvious pattern. The dots lie both above and below the 45-degree line. Regression results tell us that Average model's goodness of fit performs better than that of the Bayesian model's. The Average model performs especially well for Parm-A sessions. However, the Average model does a relatively poor job in Parm-B sessions. Although Shannon's entropy tells us that the uncertainty is similar in the parm-A sessions and parm-B sessions, fewer draws in Parm-B sessions provide less information to the subjects, and thus their elicited posteriors are noisier. The graphs of elicited probabilities' standard deviation given each predicted posterior show us that standard deviation exhibits an inversed-U shape, and such pattern in Parm-A sessions' plot has higher curvature compared with that in Parm-B sessions. This indicates that Parm-B sessions' subjects report noisier posterior beliefs at the two ending points (0,1), and thus causes the poor performance of the Average model in Parm-B sessions.

A new model called "Liner in the Ratio" (LR) that derived from subjects' feedback is introduced and tested. When using observations for signal 1, the LR model's prediction fits the elicited posterior well. Especially for Parm-A sessions, there is no inverse-S-shaped pattern in the comparison. This paper then combined the LR model's predictions for both signals in the Average way (LR(Ave)) and also the Bayesian way (LR(Bayes)). While neither LR(Ave) or LR(Bayes) did a significantly good job in both Parm-A and Parm-B sessions, LR(Ave) does fits the data well in Parm-B sessions after the outliers were dropped. The reason could be that the reduction of draws in Parm-B sessions also reduced the difficulty of calculating the ratio of black balls, and thus it facilitates LR model's goodness of fit. After dropping the outliers, it enhanced the goodness of fit for LR(Ave) model. Bayesian Multiplication model and Weighted probability Function assuming both heterogeneity and homogeneity are tested as well. The regression results do not provide clear evidence that any of the models fits the data well.

At the individual level, I divide all possible combinations of p_1 and p_2 into four cases and focused on case 1 and case 4 first. The results are mixed; most of the subjects exhibit both Bayesian and Average behaviors. While the confused behaviors did not take much proportion, the no-response behaviors took a big proportion of subjects' behaviors. The reason that explains the big proportion of no-response behaviors is highly related to LR model. When facing the signals, subjects tend to calculate the ratio of the black balls, when the ratio is close to each other in signal 1 and signal 2, subjects tend to choose to stay at their previously reported beliefs. Regressions on each individual are tested, and the results show that different subjects exhibit different behaviors. While most of the subjects can not simply be categorized into either Bayesian, Average or LR model, some of the subjects' behaviors are well explained by one of the models respectively.

Regressions of estimating the round effect (case 1 and case 4 only) indicates that, for Parm-A sessions, as each additional round accomplished, subjects are 0.436 percentage point less likely to behave like Average. This learning pattern might be because subjects eventually learned that being a Bayesian in this game could result in a higher expected payoff. But Parm-B sessions do not show us any effect from the round in the two cases. When I look at the individual's learning pattern for all cases, the convergence regression that employed from Noussair et al. (1995) shows individuals elicited belief do not converge to neither Bayesian's prediction nor Average's prediction.

Many extensions could be made from this paper. One thing we could do is to reveal the true state of the world at the end of every round so that subjects could get involved into more learning processes. Then we could study if subjects' posterior beliefs are able to converge to Bayesian predictions over time. In addition, communication was not allowed in this experiment, study how social learning interacts belief updating could be also interesting. A further step we could take is to test the impact of network structures on social learning.

Chapter 2

Motivated Beliefs, Markets, and Information Aggregation

2.1 Introduction

Substantial evidence suggests that people suffer from motivated beliefs, i.e., believing what they want to believe. Such motivation biases the acquisition and the evaluation of evidence so as to support the conclusions that people prefer (Epley and Gilovich (2016)). It is natural to ask about the impact of motivated beliefs on market performance. A few recent studies (e.g., Cueva and Iturbe-Ormaetxe (2021), Gödker et al. (2021)) explore its impact on traders' belief-updating at the individual level.

As yet, there are very few studies on how motivated beliefs affect market performance. It is difficult to address this issue with field data because beliefs, private information, and belief updating processes are largely unobserved. To fill that gap, we induce motivated beliefs and investigate their impact on asset market prices and allocations in the laboratory. Traders naturally have homegrown motivated beliefs for our common value asset that pays state-contingent payoffs; to a greater or lesser degree, we expect them to prefer the state that pays the higher payoff. We, therefore, induce polarized motivations over the different states of the world. This design enables us to clearly identify the separate impact of both sorts of motivated beliefs.

Specifically, this paper aims to answer the following two main research questions. First, how do polarized motivated beliefs affect market efficiency? Second, does trading experience attenuate or exaggerate polarized motivated beliefs? Our first hypothesis is that markets are less efficient when traders hold polarized motivated beliefs since traders' offered prices and trading strategies reflect their distorted motivated beliefs, and thus information aggregation is impaired. Our second hypothesis is that trading experience will mitigate the polarization in beliefs because interactions with other traders in the market can help traders reach a common posterior belief.

The experimental design features a control treatment with no induced motivation and a polarized induced motivation treatment. The basic setup includes human traders using a computer platform to trade a state-contingent common value asset for multiple periods. The assets in both the control and treatment sessions have the same dividend schedule. Subjects begin with a common prior that both states are equally likely.

Motivated beliefs are induced in the treatment. To do so, subjects begin by taking an intelligence quotient (IQ) test, and their IQ scores are measured.[1] It has been proven in many previous studies that most subjects have the motivated belief that they

[1]Subjects complete ten Raven's Progressive Matrices, a traditional part of an IQ test. For brevity, in this paper, as the IQ test.

did relatively well in the IQ test regardless of their actual performance.[2] Then, a group assignment device ties traders' motivations for having a high IQ test score to a state so that half the traders are motivated to prefer one state and the other half to prefer the alternative state because the preferred state signifies a good IQ test performance. In both control and treatment, traders would have the homegrown motivation for the state that pays a higher dividend. We denote this motivated belief as "Payoff motivation." The polarized motivated beliefs induced by the IQ test in the treatment are denoted as "IQ motivation." The comparison between the control and the treatment illustrates the impact of polarized motivated beliefs, which is the main focus of our study. Our design can help us decompose and compare these two motivations under different market environments.

After that, traders in all treatments follow the same protocol: each trader independently receives a private signal with a known accuracy indicating which state is the true state, and traders then start trading in the market. During each trading period, three types of beliefs are elicited from each trader: the point probability of one of the states being true, the valuation of the asset, and the anticipated ranking in trading profit. The three elicitation tasks borrow methods from Karni (2009), Becker et al. (1964), and Barron and Qu (2014), respectively, to incentivize truth-telling. Traders' beliefs are elicited at key junctures to track and study their belief-updating processes during the trading. This design could help us answer our second research question. At the end of each period, feedback pages reveal information about traders' own performance in this period.

[2] See Oprea and Yuksel (2020) and citations within.

To provide more robust evidence of our findings, we conduct an additional treatment that varies the dividend of the assets so that traders face more intense financial stakes. This new treatment differs from the initial treatment only in dividend payments. In addition, we are also concerned about traders' session experiences and the accuracy of the private signal might affect market performance. Therefore, we ran extra experimental sessions for all treatments with more precise private signals and invited subjects who had previously participated in a similar experiment.[3] The experimental design section elaborates on the details.

Our study includes several distinctive design features. First, a novel procedure induces traders' motivated beliefs with polarization. Second, a set of questions track traders' beliefs at several key junctures to provide evidence of how traders update their beliefs throughout the trading period. Third, financial stakes with different intensities and private signals with different precisions are studied to provide more robust evidence under different market environments.

The main findings of this paper suggest that polarized motivated beliefs do not seem to significantly impact market efficiency, which is measured by how far the observed prices deviate from the Rational Expectation Price (RE-price, henceforth). However, the polarized motivation impacts traders' strategies. Traders who prefer the state with a higher dividend hold significantly more assets than those who prefer the alternative state. Thus the price spread is smaller than that in markets with no polarized motivation. Traders' beliefs are polarized on point probability and asset valuation. Such

[3] We denote these traders as "experienced" traders. They either participated in one of this project's sessions or a companion project, Wang et al. (2022). The companion paper uses a nearly identical market user interface, but the main focus is the interactions between markets and traders' homegrown overconfidence.

polarization persists after traders receive the private signal and trade in the market. No evidence indicates traders' belief errors — relative to Bayesian — are mitigated after trading. In addition, the findings suggest that a more intense financial stake might improve market efficiency when traders are less experienced and hold less accurate signals. Finally, we find that IQ motivation is significantly stronger than Payoff motivation in our setting.

This paper contributes to two strands of the literature. The first contribution is to the literature that studies information aggregation. Since at least Plott and Sunder (1988), experimental economists have seen that markets may be able to aggregate private information (e.g., Choo et al. (2019), Friedman and Aoki (1992), Forsythe and Lundholm (1990)). Many factors that might affect information aggregation have received substantial attention. Numerous previous studies focused on external factors such as insiders' impact (e.g., Plott and Sunder (1982), Banks (1985), Camerer and Weigelt (1991)), the cost of acquiring information (e.g., Sunder (1992), Kraemer et al. (2006), Kübler and Weizsäcker (2004)), and the presence of futures markets (e.g., Friedman et al. (1984)). On the other hand, traders' internal belief distortions, such as optimistic bias (e.g., Ackert et al. (2008)), asymmetric learning (e.g., Kuhnen (2015)), and overconfidence (e.g., Wang et al. (2022), Magnani and Oprea (2017), Biais et al. (2005), Deaves et al. (2009), Glaser and Weber (2007), Kirchler and Maciejovsky (2002), Fellner-Röhling and Krügel (2014)), also have significant impact on traders' decision making and/or overall market performance. However, there is still much to learn about the effect of traders' internal distortions on the market. A few recent papers specifically focus on the role that motivated beliefs play in affecting traders' decision-making.

For example, Cueva and Iturbe-Ormaetxe (2021) found that buying a stock induces optimistic expectations when the stock is a loser and Gödker et al. (2021) found that individuals exhibit asymmetrical memories of outcomes from their chosen investments. To our knowledge, this is the first paper that studies the impact of polarized motivated beliefs on information aggregation.

This present paper also contributes to the growing literature on motivated beliefs. Many previous works studied individuals' belief updating processes when subjects hold motivated beliefs. While some papers (e.g., Eil and Rao (2011), Mobius et al. (2011), Grossman and Owens (2012),Oprea and Yuksel (2020)) found evidence of asymmetric updating, others (e.g., Barron (2021), Coutts (2019), Gotthard-Real (2017)), however, did not. In addition, Oprea and Yuksel (2020) found that social communication can amplify motivated beliefs. Instead of studying this point within an individual choice design, this paper provides a new perspective on motivated beliefs in market environments.

Furthermore, the experimental designs in many previous laboratory work induced motivated beliefs but did not polarize them. Eil and Rao (2011) and Oprea and Yuksel (2020) only have one of the two states reinforce all subjects' ego; similarly, subjects in Charness and Dave (2017) would only prefer the state with a higher payoff. However, people are not always motivated to prefer the same state. For example, people's beliefs regarding factual issues may be polarized due to Pro-Democrat and Pro-Republican motivations (e.g., Alesina et al. (2020), Thaler (2021), Barron et al. (2022)). In prediction markets, traders trade assets that pay dividends depending on whether a presidential candidate will win the next election. We might expect polarized motivated beliefs to impact individual trading strategies and market efficiency in markets like this. This study adapts the IQ test method from Oprea and Yuksel (2020) and develops a novel group assignment device so that subjects have polarized motivations over two states. This group assignment device can be applied by future studies focusing on motivated beliefs with polarization. In addition, financial-relevant and ego-relevant motivations have received growing attention in the literature (e.g., Coutts (2019)). This paper contributes to the literature that induces various motivated beliefs and provides evidence to show that motivation induced by IQ tests is stronger than the motivation induced by higher payoffs.

The paper proceeds as follows. Section 2.2 lays out experimental design and the implementation details. Section 2.3 presents the results. Section 2.4 concludes. Appendix B.1 offers supplementary figures and tables and Appendix B.2 provides instructions to laboratory subjects.

2.2 Experimental Design

Section 2.2.1 explains how the polarized motivated beliefs are induced and introduces the procedures for each session. Section 2.2.2 presents the two treatment environments of normal as well as intense financial stakes and extra sessions that increase the precision of the private signals. Finally, Section 2.2.3 describes the details of the implementation of the experiments.

2.2.1 Sessions Procedures and Treatment Design

All sessions follow the same procedures shown in figure 2.1. First of all, subjects need to complete a Raven's Matrices Test (IQ test). After the instructions, they play the game for several periods, including two practice periods. Each period consists of the five parts shown in figure 2.1. Then subjects are paid, and the session ends. Some details of the procedure follow.

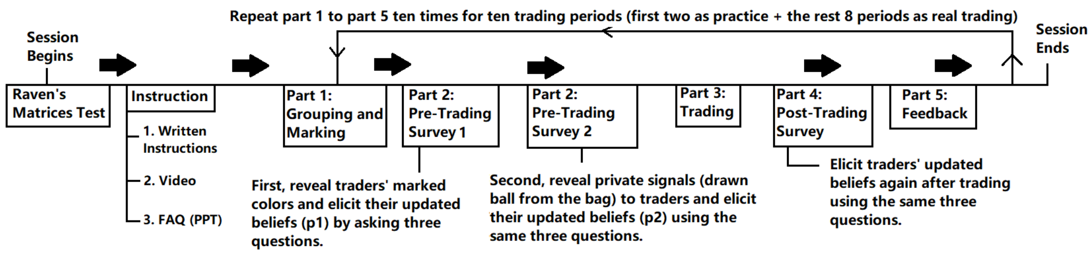


Figure 2.1: Session Timeline

Raven's Matrics Test. Subjects first need to complete a Raven's Matrics Test that contains ten questions (the difficulty ranges from relatively easy to relatively difficult) and are given 75 seconds to complete each question. Subjects are told that one of the ten questions will be selected by random and they will receive \$5 if their answer to that selected question is correct and \$0 otherwise.[4] This Raven's Matrics Test serves the purpose of inducing subjects' motivated beliefs (IQ motivation). Due to self-image motivations, most subjects will believe that they performed relatively well in this IQ test, according to many previous studies.[5]

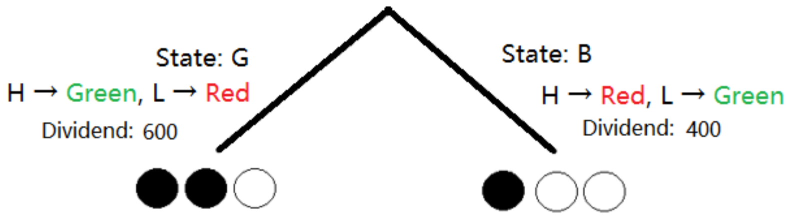
Instructions. Subjects then read the written instructions that consist of two separate files: (1) the instruction that explains the game and (2) the "Survey Questions

[4] Subjects were not informed about the purpose of the Raven's Matrices Test until they started reading the instructions. Test scores are recorded but not revealed.

[5] See Oprea and Yuksel (2020) and citations within.

Procedure," which explains the truth-telling mechanisms of the survey questions they need to answer during the game. Subjects then watch a video for visualized explanation. The experimenter then goes over a few PowerPoint slides to clear up any points of confusion from the pilot sessions. The entire instruction process takes about 40 minutes. In the post-experiment survey,[6] all the subjects reported that they understood the game well.

Traders then start the game, which consists of multiple trading periods. Each period consists of the same five parts shown in figure 2.1. The control and treatment only differ in part 1. Details of the five parts are explained below.



Note: Figure 2.2 also shows the state-contingent virtual bags that contain black and white balls, which are used to send the second private signal to traders in part 2. A black (white) ball drawn indicates the true state is more likely to be G (B).

Figure 2.2: Grouping and Marking

Part 1: Grouping and Marking. As shown in figure 2.2, there are two states of the world, Good (G) and Bad (B), with equal probability. The hypothetical asset pays 600 experimental points if the state is G and 400 if the state is B. In the treatment, the "Grouping and Marking" protocol ties traders' motivations of having a high IQ score to the two states so that some traders would prefer the true state to be G, while others would prefer it to be B. The following explains how the protocol works.

[6] One of the survey questions asks subjects "Did you find any part of the game confusing? If so, what was most confusing?". Their answers indicate they had a good understanding of the game.

Each trader is assigned to either the H or L group at the beginning of each period according to the following rules. Traders are told the rules but are not told the results of the grouping assignment.

Control session rules: In each period, half of the traders are randomly assigned to the H-group, and the other half are assigned to the L-group. In this case, because the grouping assignment is not related to their IQ test performance, traders' polarized IQ motivations are negligible.

Treatment session rules: In each period, the computer randomly sorts all 8 traders into 4 pairs (2 traders in each pair). In each pair, the trader with higher (resp. lower) rank on the ten IQ questions is assigned to the H (resp. L) group. Tied ranks are broken randomly. In this case, H-group and L-group have the same size (4 traders in each group). However, due to IQ motivation, traders are likely to believe that they were more likely assigned in the H group, indicating that they did relatively well in the IQ test.

Traders are then marked with a color according to the grouping assignment result and the true state. As shown in figure 2.2, State G marks H-group traders as green and L-group traders as red, and the color marking is reversed if the state is the B state. Traders are never told the true state but are told their colors in part 2 as their first private signal. As described in the treatment session grouping rule, green (red) traders in the treatment sessions would like to believe that the true state is G (B), which indicates they are from the H-group, signifying their IQ test performance is relatively better. By contrast, the control sessions would have no such polarized motivated beliefs since the grouping assignment is completely random.

Part 2: Two Pre-Trading Surveys.[7] At this point of a session, the true state is randomly determined but not announced. It is publicly known that each state, G or B, is equally likely to be the true state; therefore, traders' prior beliefs (denoted as p_0) that the true state is G, would be 0.5, presumably.

Pre-trading Survey 1: The computer first reveals traders their first private signal (color signal, henceforth), marked color in this period. Traders need to update their beliefs after receiving this color signal and report their updated beliefs (denote as p_1) by answering three survey questions that elicit their beliefs: (1) point belief that the true state is G, via the Karni method; (2) willingness to accept (valuation of the asset) via the Becker-DeGroot-Marschak (BDM) method; and (3) anticipated ranking in trading profits, via the Barron and Qu method.[8] Traders have 60 seconds to complete this survey page, and they will go to pre-trading survey 2 after they have completed this page. The payment formulas for the three questions appear in Appendix B.2.3, along with proof sketches that truth-telling is dominant. One concern is that the elicitation procedure might be confusing to some subjects. In addition to proving to the subjects that truth-telling is their dominant strategy, the visualized video instruction explains the mechanism using intuitive examples to aid subject understanding.[9] The post-experiment survey and the subjects' verbal feedback suggest that they understood the elicitation mechanism well.

Pre-trading Survey 2: Independently for each trader, the computer randomly

[7]There is no user interface for the "Grouping and Marking" protocol; it is done by the computer automatically. Traders go to the pre-trading surveys directly as the first user interface in each period. See Appendix B.2.1 for the user interface and the wording of survey questions.

[8]For each question, we use what seems to be the most standard elicitation method.

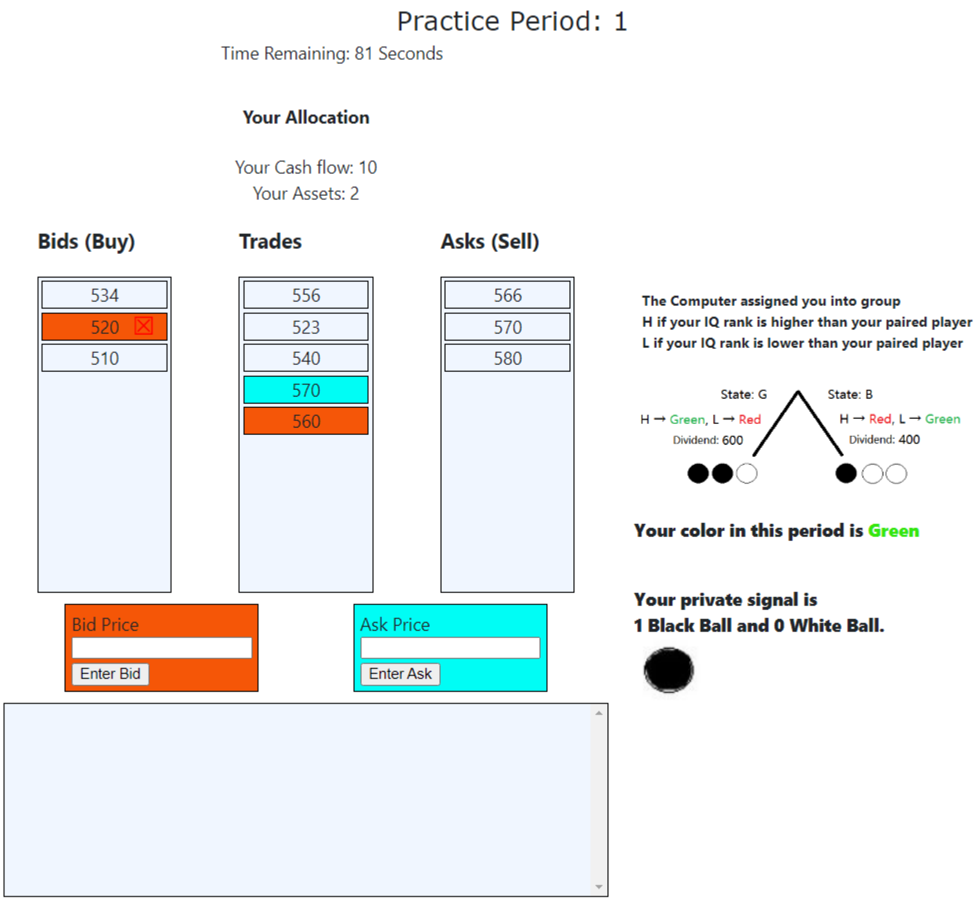
[9]See Appendix B.2.5 for an example.

draws one of the balls from the bag (shown in figure 2.2) that corresponds to the true state as their second private signal (ball signal, henceforth). Traders need to update their beliefs again and report their updated beliefs (denote as p_2) by answering the same three questions. Traders have 60 seconds to complete this second survey page.

The two pre-trading surveys help traders guess the true state and evaluate the asset before entering the market. They can also provide evidence of how traders update their beliefs given those private signals. After the two pre-trading surveys, traders go to part 3 and start trading simultaneously.

Part 3: Trading. Each trader is endowed with two units of assets that each pays the state-contingent dividends shown in figure 2.2. There is no cash budget constraint.[10] In each period, 8 traders trade for 3 minutes in a continuous double auction (CDA) market for this common value asset; see figure 2.3 for a screenshot of the trading user interface (UI). No short sales are allowed, and the maximum amount of assets one trader can hold is restricted to 8, which is half of the total units in the market. As shown in figure 2.3, bids, asks, and transaction history are public information that all traders can observe. Traders' own offers and transactions are colored as orange for buying and blue for selling. The trading prices are ordered by transaction time, with the most recent trades on the top. The UI shows traders' allocations, including the cash flow (i.e., receipts from asset units sold minus cost of units bought) and asset holdings. The right-hand side of the UI reminds the traders of the dividends, grouping rule, color marking rule, and their private signals. There is a large "Error Message" box at the bottom of the screen, and it tells traders when they make a mistake such as trying to

[10]This setting simplifies the game and can prevent some subjects from practicing no trading and holding their endowed cash to exchange for final dollar payment.



Note: Figure 2.3 shows the screenshot for the treatment sessions, where the top of the right-hand side of the UI reminds the traders that the grouping assignment is related to their IQ test performance. However, in the control sessions, the grouping assignment rule on the top of the right-hand side would show, "The computer assigned you into Group H or L by random."

Figure 2.3: User Interface for Trading

sell when they currently hold zero units of the asset.[11] Assets expire after each period, and traders do not hold the assets to the next periods.

Part 4: Post-Trading Survey. To learn how traders update their beliefs from their trading experience, they will go to the post-trading survey after the trading. The same three questions are asked to elicit their updated beliefs (denoted as p_3). In

[11] Other error messages include 1. traders cannot accept their own offers and 2. traders cannot hold more than 8 units of assets.

addition, the transaction history is recorded and displayed in this post-trading survey. All the trading prices are ordered from the highest to the lowest to help traders learn if they are, on average, buying low and selling high to adjust their trading strategy.

Part 5: Feedback. After the post-trading survey page, traders are led to the feedback pages. These pages remind traders of their final asset holdings and show them the average points they earned from answering all the survey questions and their payoff from trading. They also show the total payoff for this period, which equals the "payoff from trading" plus the "average payoff from survey questions." The feedback pages only announce state-contingent payoffs, meaning that traders are informed about their payoffs conditional on the true state (G or B). In this case, traders are not informed about the true state because revealing the true state would reveal information (via the first private signal) about traders' relative IQ test performance, which might affect traders' IQ motivation in later periods. After the feedback pages, all traders proceed together to the next period, which repeats the above five parts.

Payoff Calculations. Let n denote the amount of the final asset holdings after the trading, d be the dividend for the realized true state, R be the receipts received from sales, and C be the cost of all the units purchased from trading. Then trading profit $\pi = nd + R - C$. Each survey question also results in a payment, which is rescaled so that each question pays a similar amount of points. The average payoffs from the survey questions are about half the average trading profit. The average payoff earned from the survey questions is denoted as τ. Each trader's total payoff from one period equals the trading profit plus the average payoffs from the survey ($\pi + \tau$). All the experimental points each trader earned are summed up for their final payment for all the trading periods.

2.2.2 Treatments for Robustness

Let us denote the treatment introduced in the above subsection as the "Normal Stakes Treatment" (Treatment-Normal, henceforth) because it shares the same asset dividend (600 for state G and 400 for state B) as the control sessions. Another treatment, the "Intense Stakes Treatment" (Treatment-Intense, henceforth), is designed to be compared with treatment-normal to study how the results would change if traders face a more intense financial stake to provide more robust evidence of the findings. The only difference between treatment-normal and treatment-intense is that the dividend paid from each asset in treatment-intense is 1000 for state G and 0 for state B; all else stays the same.

The above three treatments only draw one ball from the bag corresponding to the true state, and none of the subjects had previously participated in a similar game in our laboratory. Our findings might differ if the accuracy of the second ball signal and traders' session experience vary. Therefore, we decided to run two additional sessions for each treatment. The additional sessions draw two balls with replacement as the second private signal. Therefore, let's denote the initial sessions with only one ball drawn from the bag as the "One Ball Treatment" (OBT, henceforth) and the additional sessions as the "Two Balls Treatment" (TBT, henceforth). Furthermore, all the subjects recruited for TBT sessions had participated either in OBT before or in another companion project (Wang et al. (2022)) that shared a similar experimental design and a nearly identical market user interface. Finally, for the OBT sessions, the same set of private signals, including the first and the second private signal, are pre-generated and are used for all the sessions so that all the sessions are informationally identical. However, the TBT sessions use another two sets of private signals independently generated from the same distributions. See Table 2.1 for a demonstration of the sessions.

The TBT sessions increase both signal precision and trader skill. The point is not to test the separate effects of these features, but rather to provide a rich environment for a strong robustness check.

2.2.3 Implementation Details

All production sessions were in-person sessions conducted in the LEEPS laboratory at the University of California, Santa Cruz, from October 2021 to February 2022. In total, 126 subjects were recruited across different majors to participate in 21 sessions (15 for OBT and 6 for TBT) using the ORSEE recruitment system (Greiner (2015)), see Table 2.1 for implementation details.[12] The experiments use software programmed by the oTree platform (Chen et al. (2016)). Exactly 8 subjects participated in each session.

Main Sessions		Robustness Check Sessions			
Control (OBT)	Normal (OBT)	Intense (OBT)	Control (TBT)	Normal (TBT)	Intense (TBT)
5 Sessions	5 Sessions	5 Sessions	2 Sessions	2 Sessions	2 Sessions

Table 2.1: Sessions Implementation

For each session, only the 8 real trading periods are used to calculate the

[12]Most of the subjects in TBT sessions are those who had participated in OBT sessions before, we invite them back for TBT sessions as experienced traders. Six of the subjects in TBT had participated in Wang et al. (2022). Therefore, we have $15 \times 8 + 6 = 126$ subjects.

earnings. The sum of the experimental points is converted to US dollars after the session. There is also a $12 guaranteed show-up fee.[13] Each participant earns, on average, $33 for a two-hour long session.

2.3 Results

The results begin with detecting and decomposing IQ and Payoff motivations in Section 2.3.1. Section 2.3.2 examines traders' belief updating processes throughout a trading period. Section 2.3.3 examines market efficiency and section 2.3.4 studies traders belief errors before and after trading.

2.3.1 IQ Motivation and Payoff Motivation

Recall that p_1 refers to elicited beliefs after revealing the color signal, which induces IQ motivation in the treatment. We expect green (red) traders' elicited beliefs of the probability that the true state is G to have an upward (downward) bias compared with 0.5 since they prefer G (B) to be the true state.[14]

We decompose IQ and payoff motivations to see if the induced IQ motivation is significant. It would be easier to separate these two motivations when traders are less affected by various information. Therefore, p_1 is probably the best candidate for this analysis. If traders hold no motivated beliefs, we expect them to state a probability of 0.5 after receiving the color signal since they know that both states are equally likely. However, the elicited p_1 would deviate from 0.5 in the two treatments due to motivated

[13] About half of the sessions paid $7 for the show-up fee, and the show-up fee was increased to $12 for later sessions to attract more participants due to the impact of the Covid-19.

[14] Do traders accurately assess their relative IQ performance? Appendix B.1.2 examines this question and uses correlation analysis to reject the premise that subjects accurately self-assess their relative IQ.

beliefs. Let us use $p_1 - 0.5$ to measure traders' motivated beliefs, consisting of both IQ and payoff motivations. We mainly focus on elicited data of question 1 because elicited beliefs of question 2 are likely affected by traders' risk attitudes. In addition, we assume both of the motivations take a positive value. To our knowledge, how these two motivations interact has not been previously examined. For simplicity, we assume they are additive and take the following form.

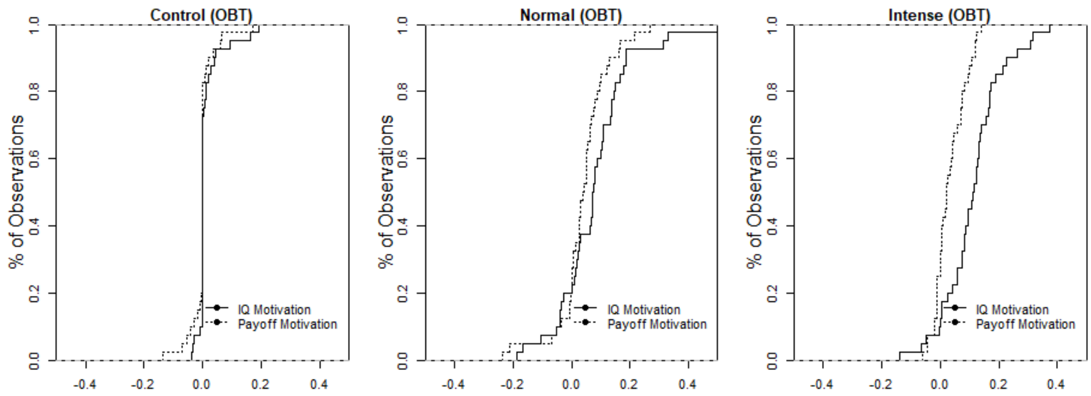
$$\begin{cases} p_1 - 0.5 = \text{IQ Motivation} + \text{ Payoff Motivation} = M_{green}, & \text{if green} \\ p_1 - 0.5 = -\text{IQ Motivation} + \text{ Payoff Motivation} = M_{red}, & \text{if red,} \end{cases} \tag{2.1}$$

Regardless of the marked color, payoff motivation would prefer state G, which pays a higher dividend. But the sign of IQ motivation would be reversed for green and red traders, indicating an opposite preference over the true state. It is likely the two motivations are heterogeneous among different individuals. We calculate the averaged motivated beliefs for each trader when they are marked as green (red) and denote it as M_{green} (M_{red}).[15] Then, we can find the two motivations for each trader using the following equations.

$$\begin{cases} \text{IQ Motivation} = (M_{green} - M_{red})/2 \\ \text{Payoff Motivation} = (M_{green} + M_{red})/2 \end{cases}$$

Figure 2.4 compares the cumulative distribution function (CDF, henceforth) of the two motivations. Consistent with our assumption, in the two treatments, both

[15]Note that all traders have been marked as green and red in each session since their colors are determined by the group assignment and the randomly determined true state.



Note: Each observation is measured at subject level. There are 40 traders in each treatment.

Figure 2.4: IQ Motivation vs. Payoff Motivation

motivations take positive values for most of the observations. In the control sessions, however, most observations take a value of zero, indicating traders' reported beliefs do not diverge much from 0.5 after receiving the color signal. The t-test shows that the mean IQ motivation is significantly greater than zero in the treatments,[16] suggesting we successfully induced traders' IQ motivation. We can also see that IQ motivation first-order-stochastically dominates Payoff motivation in the treatments. The Wilcoxon Signed-Rank test (WSR, henceforth) suggests that IQ motivation is stronger than Payoff motivation and the difference is marginally (highly) significant in normal (intense) treatment.[17]

On average, which motivation is stronger? We calculate the average of the two motivations across traders. Table 2.2 summarizes the values. Both traders' IQ and Payoff motivations are small in the control sessions, which is consistent with the graphical evidence. But in the two treatments, the average IQ motivation (Payoff

[16] The p-value of the t-test is 0.000 for both normal and intense.
[17] The p-values of the WSR test are 0.089 and 0.000 for normal and intense, respectively.

motivation) is around 0.10 (0.035), indicating the deviation of traders' elicited beliefs from 0.5 is largely due to IQ motivation.[18]

	OBT		
	Control	Normal	Intense
Average IQ Motivation	0.01 (0.04)	0.08 (0.13)	0.11 (0.10)
Average Payoff Motivation	0.00 (0.04)	0.04 (0.09)	0.03 (0.05)

Note: Standard deviations are in the parentheses.

Table 2.2: Average IQ Motivation and Payoff Motivation

The patterns remain similar in TBT sessions, the results for which are presented in Appendix B.1.3. Overall, the homegrown motivation for a higher dividend payoff is not as strong as the induced IQ motivation. Therefore, we expect traders to exhibit polarized IQ motivation over the two states in both normal and intense treatments.

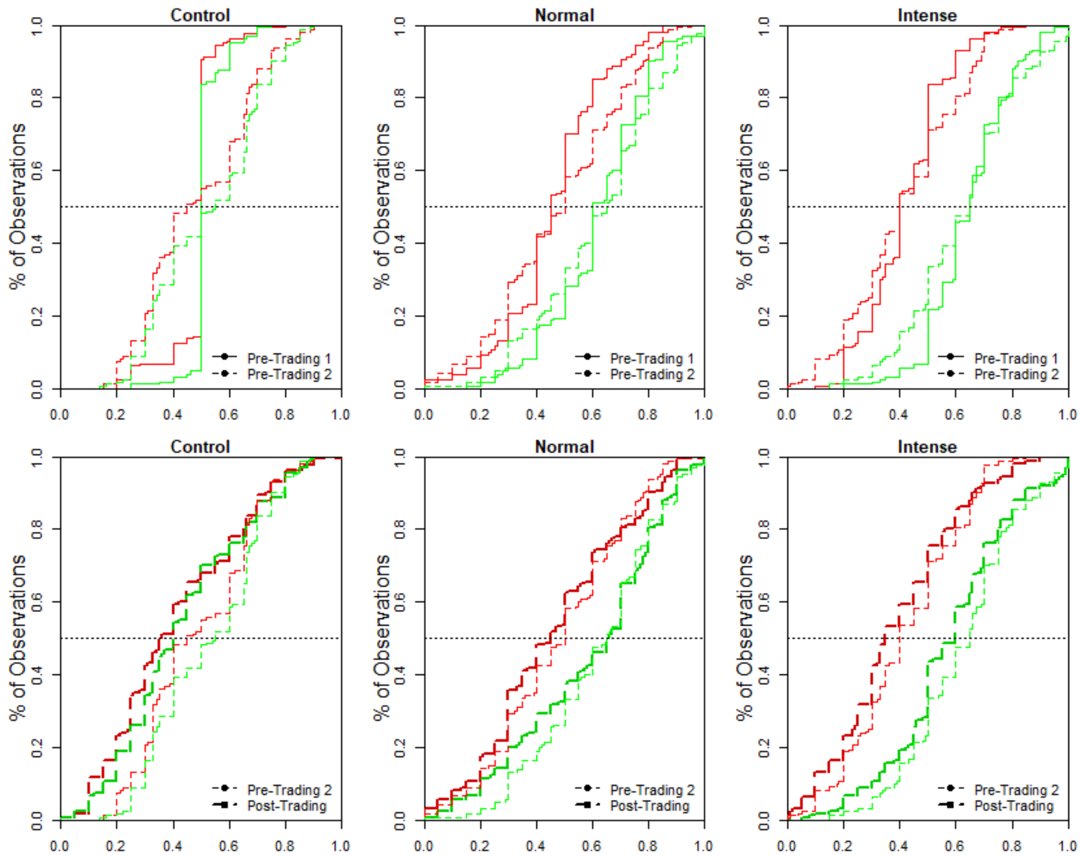
Result 2.3.1.1 *In the treatments, traders' beliefs are biased due to IQ and Payoff motivations. Overall, IQ motivation is stronger than the homegrown Payoff motivation.*

2.3.2 Beliefs at Key Junctures

This subsection illustrates the patterns of beliefs elicited in different survey pages and studies the impact of polarized motivations on traders' beliefs. Recall that p_0 denotes prior beliefs before receiving any information, p_1 denotes the elicited beliefs

[18]Readers might expect subjects to exhibit similar Payoff motivation across all three treatments. However, the data shows that subjects' average Payoff motivation is zero in control. The reason could be that the information that both states are equally likely outweighed the subjects' Payoff motivation in control. A more specialized design is needed to explore this point further. But the main focus of this paper is to induce polarized IQ motivation in the treatments.

after receiving the color signal, p_2 denotes beliefs after the ball signal, and p_3 denotes beliefs after trading.



Note: The solid line is the cumulative distribution of p_1, the dashed line is the cumulative distribution of p_2, and the thick dashed line is the cumulative distribution of p_3. Green (red) lines are beliefs from traders in the green (red) groups. The upper three graphs compare p_1 and p_2. The bottom three graphs compare p_2 and p_3.

Figure 2.5: Beliefs Updating (Question 1)

Figure 2.5 plots the CDF of the elicited beliefs of question 1 between green and red traders across three treatments (OBT sessions only). The upper three graphs compare the CDF of p_1, plotted as the solid lines, and p_2, plotted as the dashed lines. Traders' prior (p_0) that the true state is G would be 0.5 because traders know that both states are equally likely. The graph shows that, although they might have payoff motivation over state G, most traders in the control still state $p_1 = 0.5$ after receiving the color signal regardless of their colors. This is expected since the control treatment does not induce traders' IQ motivations; there is no polarization. In the normal and the intense treatments, however, we can see that green traders, whose IQ motivation prefers the good state, have an updated p_1 first-order-stochastically dominates the p_1 of the red traders. According to the WSR test, the gap between the green and red CDFs is statistically significant,[19] indicating traders with polarized motivation tend to report a belief in favor of their preferred state.

There is little gap between dashed lines of p_2 in control, indicating traders report similar updated beliefs after receiving the ball signal. By contrast, in the two treatments, the gap between green and red CDFs of p_2 is still statistically significant,[20] suggesting that the ball signal does not eliminate the belief divergence between green and red traders.[21] We also found motivated traders in the two treatments respond to the ball signals asymmetrically: they react more strongly to the ball in line with their IQ motivations (e.g., green (red) traders react more (less) to the black ball than to the white ball). This finding is consistent with many previous literature, and the detailed analysis is provided in Appendix B.1.20.

The bottom three graphs in figure 2.5 compare the CDFs of p_2 and p_3 (plotted as the thick dashed lines). After the trading, the CDFs of the green and red traders in the control sessions almost overlap, suggesting traders share similar posteriors. However,

[19]The p-value of the test is 0.000 for both normal and intense.

[20]Again, the p-values of the WSR test are 0.000 for both normal and intense.

[21]The gaps in the intense treatment appear larger than those in the normal treatment. This result is in line with some of the previous studies (See Charness and Dave (2017), for example.), and the reason could be that an intense financial stake affects traders' expectation of the true state. Thus, it further pushes green traders' beliefs away from red traders.

trading does not eliminate the belief divergence in the two treatments. Motivated traders' beliefs remain polarized after trading.

Traders' willingness to accept (WTA) is positively correlated with their point probability beliefs of question 1. Therefore, the pattern of their WTA provides similar results and are provided in Appendix B.1.4.

Despite the above divergence, green and red traders exhibit no difference in their beliefs of question 3 across all three treatments.[22] Because their profit-making depends more on their trading strategies in the market, the first two private signals do not provide much helpful information to help traders guess their rankings accurately. Thus there is no clear difference between green and red traders after receiving the private signals. In addition, trading does not distinguish between green and red traders' beliefs across treatments. This might be because the true state is not revealed at the end of the periods, and without this information, it is hard for traders to guess their relative trading performance.

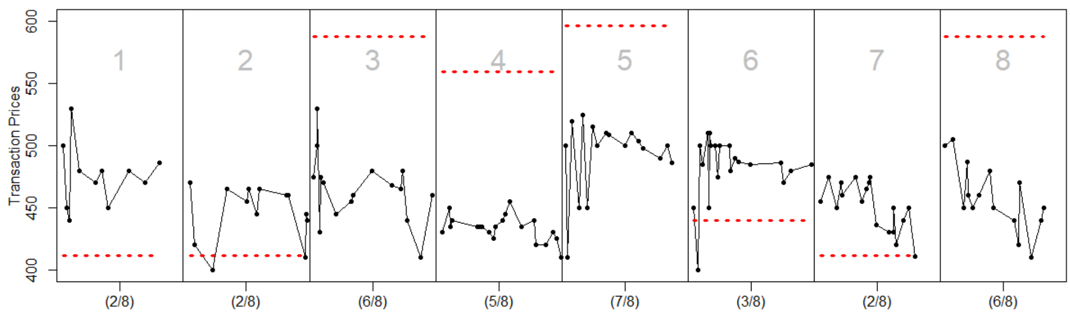
The graphs of TBT sessions are provided in Appendix B.1.6, B.1.7, and B.1.8. They provide similar patterns, suggesting the finding is robust. In addition, readers might be interested in seeing analysis using subject average as the measure. Appendix B.1.23 replicates this subsection's study but takes the average of each trader's beliefs in one session. The patterns are largely the same.

Result 2.3.2.1 *Induced polarized motivations cause a significant divergence in traders' beliefs and in asset valuation. This divergence persists after receiving a private signal and trading in the market.*

[22] See graphs in Figure B7.

2.3.3 Market Efficiency and Final Asset Allocations

This subsection analyzes how do polarized motivations impact market efficiency and trading strategy. Figure 2.6 plots the price dynamics of one control session under OBT. Numbers in gray indicate the period numbers of the session. The observed transaction prices are the black dots connected by the solid lines. Rational Expectation Prices (RE-prices, henceforth), the Bayesian prediction of the asset value conditional on all available information in the market, are plotted as red dashed lines.



Note: The transaction prices (black dots) are ordered from the first transaction (left) to the last transaction (right) within each period. Y-axis refers to the range of the transaction prices. The numbers inside the parentheses along the x-axis are the total black balls (out of 8 balls) drawn in that period.

Figure 2.6: Transaction Prices in OBT (One session Data)

The graph shows that only period seven exhibits a clear convergence towards the RE-price, but not so much in the other periods. The graphs of other sessions in OBT are provided in Appendix (See section B.1.12, B.1.13, and B.1.13), and the patterns are roughly similar. The pattern in OBT is not surprising since the ball signal is not precise, with only one ball drawn from the bag. Under TBT sessions where we provide more precise private signals, there are relatively more periods that exhibit good convergence similar to period seven in Figure 2.6 (See section B.1.15).

The regression method from Noussair et al. (1995) provides a more rigorous analysis of the convergence trend of the data. The model, shown as equation (2.2), assumes the dependent variable exhibits a convergence pattern over time within each experimental session and approaches some value asymptotically.

$$\begin{aligned} y_{jt} = &\sum_{j=1}^{5} \beta_{Cj} D_{Cj} \mathbb{1}_C \frac{1}{t} + \sum_{j=1}^{5} \beta_{Nj} D_{Nj} \mathbb{1}_N \frac{1}{t} + \sum_{j=1}^{5} \beta_{Ij} D_{Ij} \mathbb{1}_I \frac{1}{t} + \\ &+ \beta_C \frac{t-1}{t} + \beta_N \mathbb{1}_N \frac{t-1}{t} + \beta_I \mathbb{1}_I \frac{t-1}{t} + u \end{aligned} \tag{2.2}$$

The dependent variable y takes two different values in this analysis: "The price error," measured by the absolute value of the deviation of actual price from RE-price in each period, and "Spread," measured by the difference between the last best ask and the last best bid. We have j indicates the particular experiment sessions for each treatment,[23] D_{Cj} (D_{Nj} and D_{Ij}) is a dummy variable for sessions of control (normal and intense), and $\mathbb{1}_C$ ($\mathbb{1}_N$ and $\mathbb{1}_I$) is the indicator function for treatment control (normal and intense). As the number of periods t gets large, the upper three session-dependent terms vanish asymptotically because $\frac{1}{t}$ gets small. In contrast, the bottom session-independent permanent term $\frac{t-1}{t}$ converges to 1.0. Therefore, the bottom three terms are the main focus of the regression. β_C measures the asymptote of the dependent variable in control, and β_N (β_I) measures the relative difference from normal (intense) treatment.

Table 2.3 shows the estimates for the permanent terms. The "actual price" in the first column equals the last transaction price each period, while the "actual price"

[23] Equation (2.2) refers to the regression for OBT sessions. The equation for TBT sessions only changes the maximum value of j to 2 because TBT only has two sessions for each treatment.

VARIABLES	(1) P-Error (Last Price)	(2) P-Error (Last Two Prices)	(3) Spread
β_C	46.25***	48.30***	21.41***
	(3.500)	(2.803)	(4.456)
β_N	-7.101	-8.796	-12.50**
	(5.683)	(5.089)	(5.104)
β_I	-14.78**	-17.77**	-9.007*
	(6.810)	(6.291)	(4.717)
Observations	120	120	120
R-squared	0.765	0.785	0.745

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: "P-Error" refers to the price error. Because the dividend range varies in different treatments, all y variables are rescaled to 100 for consistency. Regressions use data from OBT sessions. The regressions cluster standard errors at session level.

Table 2.3: Regressions of Convergence

in the middle column equals the average of the last two transaction prices. The last column replaces the dependent variable price error with "Spread."

If the market is efficient, we would expect "The price error" to converge to zero over time. The first column in Table 2.3 shows that, according to β_C, the price error in OBT converge to a value that is significantly higher than zero, which is consistent with the patterns shown in Figure 2.6. This suggests that the markets are not fully efficient.[24] Estimated β_N measures the difference brought by IQ motivation, and the estimated β_N indicates traders' polarized preferences over the two states do not affect

[24] The markets are not fully efficient in control sessions is probably because the private signal is not precise enough. In the companion paper (Wang et al. (2022)), where we provide private signals with higher precision, we observe much clearer price convergence in similar markets. However, we want to avoid the private signal being too precise in this paper since we need traders to believe they are more likely in the H-group to maintain polarized motivated beliefs. Learning from the signal might impair IQ motivation if the precision is too strong. After discussing with other researchers, we decided to use private signals with the current precision.

price convergence much. However, the estimated β_I suggests that an intense financial stake might improve market efficiency. The middle column shows that the result remains the same if we use the average of the last two prices. Appendix B.1.16 replicates this analysis using data from TBT sessions. The estimations suggest that the results are largely robust. Polarized motivated beliefs do not have a significant impact on market efficiency.[25]

Equation (2.2) assumes the dependent variable exhibits a convergence pattern over time. We also adapt two alternative measures of information aggregation from Page and Siemroth (2021). The first measure captures the fraction of the actual number of private signals (e.g., balls drawn from a bag) that would account for the observed actual prices. The second measure captures how close the observed prices are to full information prices. These two measures tell a similar story (See Appendix B.1.22).

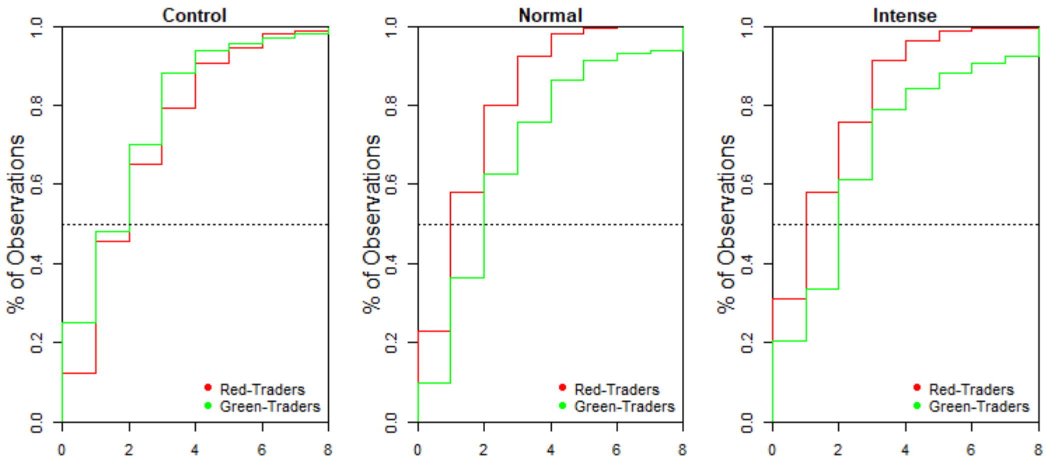
Result 2.3.3.1 *Polarized motivated beliefs do not have a significant impact on information aggregation. Increasing the financial stakes improves market efficiency when traders are less experienced and hold less precise signals.*

The last column in Table 2.3 studies price spread. The relative difference measured by β_N and β_I indicates that the spread is significantly closer to zero if traders hold polarized motivations.[26] A higher demand might lead to a narrower price spread in the financial market. Therefore, the lower spread in the two treatments might suggest a higher demand for the assets. The following analysis of final asset allocations supports

[25] We do not see that an intense financial stake improves market efficiency under TBT sessions. However, the hypothesis test on whether the two estimated β_C differ between OBT and TBT (i.e., $H_0 : \beta_C(\text{OBT}) = \beta_C(\text{TBT})$) provides a p-value of 0.001. This indicates traders' session experiences (perhaps more precise private signals) improve the overall market efficiency across all three treatments.

[26] See graphical evidence in Figure B16.

this expectation.



Note: The above graph shows the CDF of traders' shareholdings in OBT only. The horizontal dashed line indicates the median of the observations. Since the maximum amount of assets one trader can purchase in each period is limited to 8 shares, the x-axis takes values from 0 to 8. The y-axis is the proportion of observations.

Figure 2.7: Share Holdings (OBT)

Figure 2.7 plots the CDF of the final shareholdings between green and red traders in OBT. The left graph shows that green and red traders have similar shareholdings after trading in the control. However, in the normal and intense treatments, the CDF of green traders' shareholdings first-order-stochastically dominates the CDF of red traders' shareholdings. The result of the WSR test indicates that the difference between the green and the red CDFs is statistically significant.[27] This suggests that when traders have polarized motivations, those who prefer the state is G (B), tend to buy significantly more (less) assets and thus increases the demand of the assets.

The asset allocations in TBT sessions provide similar patterns, and the graphs are presented in appendix B.1.11. Appendix B.1.24 provides the same analysis but using

[27]The p-value of the test is 0.000 for both normal and intense.

subject average measure, and the patterns remain largely the same.

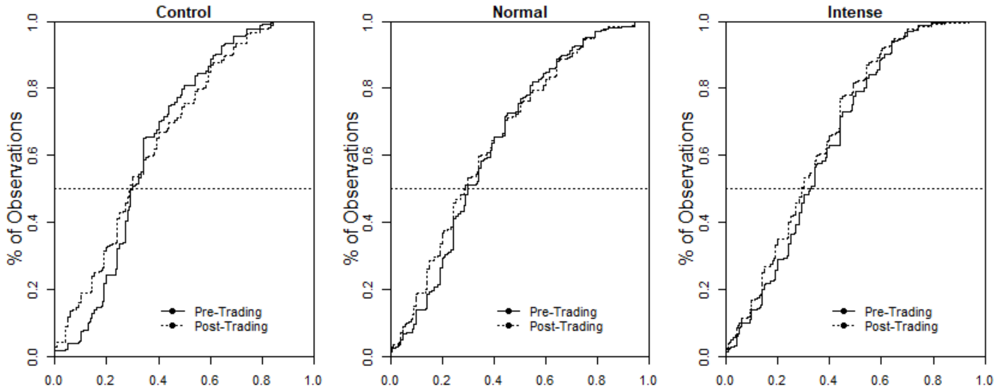
Result 2.3.3.2 *Polarized motivated beliefs significantly impact traders' final shareholdings, with traders motivated to prefer the good state hold significantly more assets at the end of trading. Such impact on shareholdings also reduces price spread when traders have less precise signals.*

2.3.4 Belief Errors

The current subsection examines whether trading attenuates the deviation from Bayesian and the role of polarized motivations. First, the belief errors are measured by the absolute difference between elicited beliefs (p_2 and p_3) and the perfect Bayesian prediction, p^*.[28] In other words, $|p_2 - p^*|$ and $|p_3 - p^*|$ is calculated for each trader, and the former measures the belief errors before trading, while the latter measures the errors after trading. Comparing the CDFs of the two belief errors could provide evidence for whether trading reduces the errors.

Figure 2.8 only provides graphs of belief errors for question 1.[29] As illustrated in section 2.3.3, OBT sessions do not have a good price convergence in all treatments; thus, the transaction prices do not provide much information regarding the true state. Therefore, we see little gap between the CDF of the post-trading errors (plotted as the dashed line) and the CDF of the pre-trading errors (plotted as the solid line) across treatments. The WSR test shows the gap is not statistically significant, suggesting there is no clear reduction of belief errors after trading.

[28] Bayesian prediction of the probability that the true state is G conditional on all market information.
[29] The graphs for question 2 and question 3 are provided in Appendix B.1.18.



Note: Belief Errors (x-axis) in this figure refer to the absolute difference between the elicited beliefs and the Bayesian prediction. The solid (dashed) lines are the CDF of the belief errors before (after) the trading. Belief Errors without taking the absolute difference are provided in Appendix B.1.19.

Figure 2.8: Belief Errors (Question 1)

However, graphs from TBT sessions, presented in Appendix B.1.17, illustrate a different pattern. In control sessions, the post-trading errors have a clear reduction relative to the pre-trading errors. The median of the belief errors shifts significantly to the left after trading.[30] In the normal treatment, however, the gap between the pre-trading and the post-trading CDFs is slightly smaller than the gap in the control sessions, and it is not statistically significant,[31] indicating that when traders hold polarized motivations, the belief errors are less reduced after trading. The pattern in intense sessions is somewhat similar to that in control sessions, suggesting an increased financial stake helps traders, even if they hold polarized motivations, become closer to perfect Bayesian after the trading. The patterns in TBT sessions suggest that traders' session experiences (perhaps a more precise private signal) can impact traders' belief updating relative to Bayesian. Given that we have limited observations in TBT sessions, we do

[30]Under TBT, the p-value of WSR test equals to 0.005 for control.

[31]Under TBT, the p-value of WSR test equals to 0.435 and 0.000 for normal and intense, respectively.

not draw solid conclusions.

Similar as in previous subsections, we also replicate the analysis of this subsection using subject average as the measure and the graphs are provided in Appendix B.1.25. The patterns remain largely the same.

Result 2.3.4.1 *In both control and treatments, traders' belief errors—relative to Bayesian—are not mitigated by market experience.*

2.4 Conclusion

Previous works on motivated beliefs mainly focus on decision-making at the individual level, while works on information aggregation haven't examined much on the impact of motivated beliefs. This study is among the first to combine the two focuses. We design a laboratory market with a common value financial asset that pays state-contingent payoffs. The design then polarizes traders' motivations over the two states and studies the impact of polarized motivated beliefs on market efficiency.

The key findings suggest that polarized motivated beliefs do not significantly impact market efficiency. However, traders' polarized motivated beliefs persist even after receiving private signals and trading in the market. Polarized motivated beliefs also significantly impact price spread and traders' asset holdings. We do not find evidence that traders' belief errors — relative to Bayesian — are mitigated after trading. In addition, an intense financial stake only improves market efficiency when the private signal is less precise. This study also found subjects' motivated beliefs induced by IQ tests are stronger than the homegrown motivation for higher payoffs. Our findings remain robust under market environments with increased signal precision and trader skill.

Our findings might interest policymakers and scholars outside of academics. Motivated beliefs polarize public opinions through social media and thus create many societal issues. How to construct a policy that helps improve market efficiency and societal stability is still a challenging task. Of course, one of the limitations of this paper is that it studies motivated beliefs in an abstract laboratory environment. By adapting methods from previous works, this paper uses IQ test performance to induce the subject's motivations. To provide insight on real-world events and markets, subsequent work can further explore more complex situations. Motivated beliefs in real life regarding politics and religion might be stronger than what was induced here. Although some field and online experiments have started using controversial real-life political events in their studies, there is still much to explore regarding new methods to induce subjects' motivations in the laboratory.

Chapter 3

Overconfidence and Market Performance

3.1 Introduction

There is a widespread impression that professional traders tend to be overconfident about their own relative ability and about the accuracy of their beliefs. If so, how would such overconfidence affect their personal performance in asset markets? How would it affect the overall performance of financial markets? But perhaps traders are not actually so overconfident. At least since Alchian (1950) some economists have argued that market participants who deviate from rationality will suffer from lower payoffs, and they either will learn to act more rationally or else will be displaced by more rational agents. Do markets indeed discipline traders and reduce their overconfidence?

To better understand how overconfidence interacts with asset market performance, we build a new laboratory asset market. Here human participants receive private information and then trade an asset that pays a single state-contingent dividend. We consider different information environments in which the private information has homogeneous precision, either high or low, and also consider heterogeneous environments in which the precision distribution either is explicitly announced or remains ambiguous. To assess overconfidence (or underconfidence), we elicit traders' beliefs about the state, the asset value, and their relative profitability. We do so right after traders receive their private information prior to trade, and again right after the asset market closes.

We find that overconfidence is a semi-coherent personal trait: measurement via belief elicitation is highly correlated with measurement via value elicitation, but not with measurement via anticipated profit rank. We find that the markets do a good job of disseminating private information, and that price efficiency is enhanced by trader experience and by unambiguous precision distribution. Overconfidence reduces price efficiency in some environments, but (perhaps surprisingly) tends to increase it in other environments. We find considerable persistence in traders' overconfidence, and little evidence that it is mitigated by market participation.

Our experiment draws on, and contributes to, several strands of literature. The impact of traders' overconfidence on market performance has been studied in the field since Odean (1998), who showed that overconfident traders have higher trading volume and hold underdiversified portfolios. Caballé and Sákovics (2003) find that public overconfidence can increase trading volume or liquidity. Later laboratory as well as field experiments find increased trading volume arising from calibration-based overconfidence (Deaves et al. (2009), Biais et al. (2005)), the better-than-average effect (Glaser and Weber (2007)), and misperceived signal reliability (Fellner-Röhling and

Krügel (2014)). Overconfidence also affects asset price and market bubbles (e.g., Aragón and Roulund (2020), Michailova and Schmidt (2016)). A few recent papers (e.g., Bregu (2020), Meier and De Mello (2020), Ida and Okui (2020)) find that overconfidence declines when participants are given feedback, but other papers (e.g., Hoffman and Burks (2017), Huffman et al. (2019)) find more persistent overconfidence.

Existing literature seems largely to neglect heterogeneous precision and ambiguity regarding precision. Partial exceptions include Lunawat (2021), who publicly announce the average of subjects' pre-trading dividend forecast. Kirchler and Maciejovsky (2002) vary the precision of the public signals and find that participants are not generally prone to overconfidence. Barron and Qu (2014) include a high asymmetry treatment in which only half of the traders receive private signal. Our experiment seems to be the first to systematically examine the market impact of over- and underconfidence given varying signal precision and also given ambiguity regarding relative precision.

A separate strand of literature is concerned with the measurement of overconfidence. Moore and Healy (2008), Hilton et al. (2011), and Glaser et al. (2013)) recognize four sorts of overconfidence. (i) Judgmental overconfidence refers to overestimating the precision of one's judgment. (ii) Self-enhancement bias, also known as the better-than-average effect, refers to an unrealistically high estimate of one's own rank or relative position. (iii) Overestimating the quality of one's absolute performance. (iv) Over-optimism regarding societal risks.

Judgemental overconfidence (i) is the sort most commonly examined in the literature. Typically (e.g., in Meier and De Mello (2020), Glaser et al. (2013), Biais et al. (2005)) the measurement task is to elicit a 90% confidence interval for each of several numerical general knowledge questions (e.g., how long is the Nile river?); the subject is deemed overconfident if significantly more than 10% of correct answers lie outside those confidence intervals. Two of our own survey questions also concern judgemental overconfidence, but use direct measures that connect to the asset market. We will say that a subject is overconfident (resp. underconfident) if their elicited probabilities or valuations reflect more precise (resp. less precise) private information than she actually has. We will also include a direct measure of self-enhancement bias that connects directly to our asset market.

A companion paper (Wang (2022)) uses similar procedures to elicit traders' beliefs and their market impact. The present paper focuses on homegrown overconfidence, while the companion paper focuses on induced polarized motivated beliefs.

The present paper unfolds as follows. Section 3.2 collects relevant standard theoretical material. Section 3.3 lays out the laboratory procedures and experimental design, and lists the specific hypotheses that we will test. Section 3.4 presents data summaries and hypothesis test results. Section 3.5 concludes. Appendix C.1 offers supplementary figures and tables and Appendix C.2 provides instructions to laboratory subjects.

3.2 Theoretical Considerations

Consider a world with two equally likely states, {G, B}, and imperfect signals about the true state. Low precision signals s_L are independent Bernoulli trials that indicate the true state with probability $q_L \in (0.5, 1)$, and high precision signals are similar but with probability $q_H \in (q_L, 1)$. Bayes theorem tells us that, after observing N_L low precision signals of which L indicate state G and N_H signals of which H indicate G, the posterior probability that the true state is G is

$$P = \frac{q_L^L \cdot (1-q_L)^{N_L-L} \cdot q_H^H \cdot (1-q_H)^{N_H-H}}{q_L^L \cdot (1-q_L)^{N_L-L} \cdot q_H^H \cdot (1-q_H)^{N_H-H} + (1-q_L)^L \cdot q_L^{N_L-L} \cdot (1-q_H)^H \cdot q_H^{N_H-H}}. \quad (3.1)$$

Figure 3.1 illustrates with an example used in the experiment. The left side presents s_L in terms of two bags that each contain 5 balls. If the true state is G (resp. B) then the decision maker sees only balls drawn randomly (with replacement) from the "Good bag" that contains 3 black balls and 2 white balls (resp. from the "Bad bag" with 2 black balls and 3 white balls). The decision maker does not see the bag, only the balls drawn from it. Thus $q_L = \frac{3}{3+2} = 0.6$ is the low precision signal accuracy, i.e., the probability that a black (resp. white) ball correctly indicates state G (resp B). The right side of Figure 3.1 similarly illustrates that the high precision signal s_H has accuracy $q_H = 0.8$. For example, consider two low precision signals, so $N_L = 2, N_H = 0$ in equation 3.1. The equation tells us that the posterior probability of state G in this example is $\frac{0.6^2 \cdot 1 \cdot 1 \cdot 1}{0.6^2+0.4^2} \approx .692 \approx 70\%$ if both balls are black (so $L = 2$), versus about 30% if both balls are white or exactly 50% if one ball is black and the other is white.

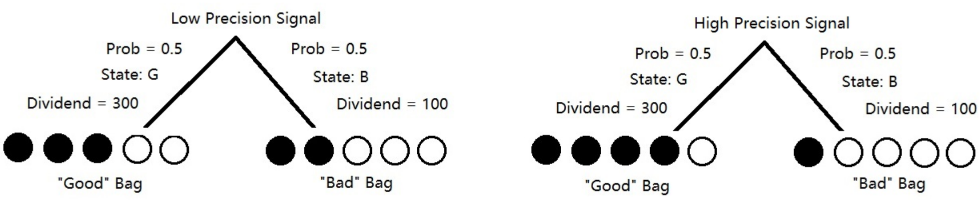


Figure 3.1: Private Signals

Of course, human subjects can't be expected to always implement equation (3.1) exactly. Recall the first notion of overconfidence in the literature is that private information is treated as if it were more precise than it actually is. To formalize that notion in our setting, note that increasing a uniform signal precision q in (3.1) pushes the posterior probability further away from the prior probability 0.5. Also, recall the signum function $sgn(y) = +1$ if $y > 0$, $=0$ if $y = 0$, and $=$-1 if $y < 0$. Suppose that a subject reports subjective probability p of state G when, according to equation (3.1), the true probability is P. Our *index of overconfidence* is then defined as

$$x = (p - P)sgn(P - 0.5). \tag{3.2}$$

That is, the report p is overconfident ($x > 0$) to the extent that it is further away than P from the prior probability 0.5 of state G, and it is underconfident ($x < 0$) to the extent that it lies closer to 0.5 than does the Bayesian posterior probability P. An interesting case is when $P = 0.5$ (the signal realization is one black ball and one white ball) and the report p deviates from P. In this case, since traders do not know which state is the true state, any deviation from P is categorized as overconfident and we set x to take a positive value, $|p - P|$.[1]

We would like to measure subjects' overconfidence x for both pre and post-trading but, unfortunately, it is impractical to do so. It turns out that in the vast majority of realizations, the fully-aggregated probability P given all private signals is very close to one of the two end points (as in Figure 3.6), and so leaves no room to observe overconfidence in post-trade elicited beliefs. Therefore, we only report x using traders' pre-trade beliefs. P in equation 3.2 is the Bayesian posterior, given an individual

[1] If p and P lie on opposite sides of 0.5, then $x < 0$. Later we will redefine such a report as "confused" rather than underconfident.

trader's private signal.

3.2.1 Information Aggregation

An asset market is informationally efficient to the extent that it aggregates private information into the asset price. For example, suppose that each unit of an asset pays $d_G = 300$ in state G and $d_B = 100$ in state B. Suppose also that the state is revealed only after the market closes, but that each trader receives private information before the market opens. Then the market is informationally efficient to the extent that actual transactions prices converge to the fully-aggregated or rational expectations price

$$V^* = Pd_G + (1-P)d_B = 100 + 200P, \tag{3.3}$$

where P = the Bayesian posterior probability from equation (3.1) conditional on all private signals, with N_L = number of low precision signals received by all market participants, L = the total number of such signals indicating state G (e.g., the total number of black balls from low precision bags), and with N_H and H similarly defined for high precision signals.

For example, suppose that (as in period 13 of Figure 3.4.2 below) there are four participants who each receive two low precision ($q_L = 0.6$) signals and 5 of those 8 signals indicate state G, and four other participants who each receive two high precision signals ($q_H = 0.8$), of which 2 indicate G. Then $P = \frac{0.6^5 \cdot 0.4^3 \cdot 0.8^2 \cdot 0.2^6}{0.6^5 \cdot 0.4^3 \cdot 0.8^2 \cdot 0.2^6 + 0.4^5 \cdot 0.6^3 \cdot 0.2^2 \cdot 0.8^6} \approx 0.0087 \approx 1\%$ and $V^* \approx 102$. Although low precision participants have mostly misleading information, in this example at least two of the high precision participants see two white balls and so have pre-trade posterior probability of G of $\frac{0.2^2}{0.8^2+0.2^2} \approx 0.0588 \approx 6\%$. With full aggregation, those two (or possibly three) pieces of high-quality private information far outweigh the misleading low precision information.

It is difficult to model the actual process by which self-interested asset market traders might aggregate their private information (see Copeland and Friedman (1991) for an early attempt), but the general idea is that offer prices and trade prices may reveal some of the market participants' private information to the other participants. As a practical matter, aggregation will typically be less than complete, so it is useful to have measures of incomplete aggregation, or informational inefficiency. Perhaps the most direct measures are in terms of deviations of actual prices ν from the true value V^* given in equation (3.3). The mean absolute deviation (or, alternatively, the root mean squared deviation) of ν from V^* is a natural inefficiency metric that we will use in empirical work. We will also use an alternative metric proposed by Page and Siemroth (2021). Their idea is to find the fraction $\lambda \in [0,1]$ of the actual number of private signals (e.g., balls drawn from a bag) that would account for the observed precision of actual prices; see Appendix C.1.11 for details on our implementation of λ.

Signed deviations from V^* can also be used as an index of overconfidence when traders' asset valuations v are elicited instead of probabilities:

$$x = (v - V^*)sgn(V^* - 200). \tag{3.4}$$

Similar to equation 3.2, the report v is over(under-)confident when $x > 0$ ($x < 0$). The fully-aggregated price is again usually very close to one of the two end points, so again we only compute x for pre-trade elicited valuations, where V^* in equation 3.4 is the true value conditional on an individual trader's private signal.

3.2.2 Trading strategies and overconfidence

A *market format* defines how market participants (traders) can make offers to trade and how those offers are processed to create actual trades (transactions). Our experiment will use the format known as the continuous double auction (CDA), variants of which are used in most modern financial markets (Friedman and Rust (1993)).

In a CDA, there is a known time interval (3 minutes in our experiment) in which the market is open. During this time, each trader is free to submit (or cancel or replace) a publicly observable limit order to buy (a bid) and to sell (an ask). In our experiment, a bid at limit price y is an offer to buy a single asset unit at the lowest ask price currently in the order book, but if all such prices exceed y, then the bid is placed in the order book. Likewise, an ask at limit price z will execute immediately at the highest bid price in the current order book if it is at least z, but if no such bids are present then the new ask enters the order book. The trader who places a new order that executes immediately is called the (price) "taker" and the counterparty (whose order was resting in the orderbook) is called the (market) "maker."

For our asset market under any format, we can write trading profits as $\pi = nd + R - C$, where n is the number of asset units held when trading closes, d is the dividend for the realized true state, R is the revenue from selling asset units, and C is the cost of all the units purchased by a given trader. For the CDA format we can take a deeper look at the sources of trading profits. Makers earn larger profits (via increasing

$R-C$) to the extent that they (a) have a larger positive spread, i.e., ask price minus bid price, and (b) attract greater trading volume. Takers earn larger profits (via increasing the expected value of $d - C$ for asset purchases and of $R - d$ for asset sales) to the extent that they buy at prices below V^* and sell at prices above V^*, and have larger such trading volume.

We will address several open questions concerning the connection between trader profit and overconfidence. The highest volume takers might be those who are most confident about their estimates of P and V^*. But if excessive, such confidence might lead them to trade at less favorable prices. Moreover, both makers and takers might be more active (and perhaps less profitable) when they are more overconfident about their innate trading ability. Before formulating testable hypotheses on these and other matters, we lay out the structure of our laboratory experiment.

3.3 Experimental Design

3.3.1 Procedures

Each session follows the same timeline in Figure 3.2. First the human subjects receive instructions. Then they play the game for several periods, receive their payments, and the session ends. Key details follow.

Instructions. Each session begins with all subjects reading two instruction documents, one that explains the game, and another that explains why truthtelling is a dominant strategy in responding to the survey questions that they will answer. Appendices C.2.2, C.2.3, and C.2.4 include copies of these documents. Next, subjects watch a video that

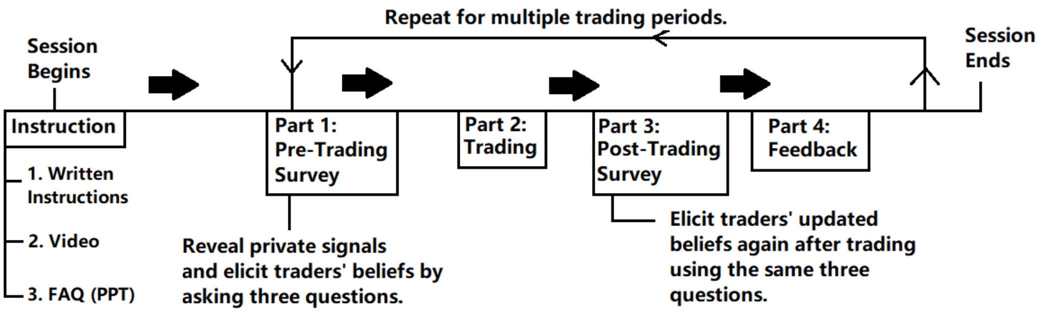


Figure 3.2: Session Timeline

features animations of the user interface. The experimenter then presents PowerPoint slides with answers to questions frequently asked in pilot sessions, and subjects are encouraged to ask their own clarifying questions. In post-experiment surveys, all subjects reported that they understood all tasks.

After instruction, subjects complete two unpaid practice periods and 14 paid periods. As the timeline indicates, each period has the following four parts.

Part 1: Pre-Trading Survey. Subjects see two balls drawn with replacement from an unseen bag related to the true state, as detailed in Section 3.3.2 below. Then a three question survey elicits traders' beliefs about the probability of state G (Question 1, using the Karni (2009) mechanism), about their willingness to accept for an asset unit (Q2, using the BDM mechanism, cf. Becker et al. (1964)), and about their relative trading profit in the current period (Q3, using the Barron and Qu (2014) mechanism). The exact payment formulas for Q1-Q3 appear in Appendix C.2.4, along with proof sketches that truth-telling is dominant.

Part 2: Trading. Each trader is endowed each period with two units of an asset that pays the per-unit liquidating dividend of $d_G = 300$ in state G and $d_B = 100$ in equally likely state B. There is no cash budget constraint. In each period, 8 human subjects trade for 3 minutes in a continuous double auction (CDA) market for this common value asset; see Figure 3.3 for a screenshot of the trading user interface (UI). No short sales are allowed, and traders are not allowed to hold more than 8 asset units, i.e., half of market supply. As shown in Figure 3.3, current bids and asks, and transaction price history, are public information seen by all market participants. Traders' own offers and transactions are colored green for buying and blue for selling. Transaction prices so far are shown in chronological order with the most recent trades on top. The UI shows traders' current holding of cash (including receipts from asset units sold minus cost of units bought) and the asset. The right-hand side of the UI reminds the traders of the state-contingent dividends, and their private signals. There is a large "Error Message" box at the bottom of the screen; it alerts subjects when an order is rejected, e.g., an ask by a trader currently holding zero units.[2] Asset units expire each period after paying a state-dependent liquidating dividend d; they do not carry over into the next period.

Part 3: Post-Trading Survey. To learn how traders update their beliefs from their trading experience, we conduct a post-trading survey. The same three questions — Q1, Q2 and Q3 as in Part 1— are asked again, but this time while the subject sees the transaction history. To help them improve their trading strategies, subjects also see the color-coded transaction prices sorted from highest to lowest, making transparent their success (or lack thereof) in buying low and selling high.

Part 4: Feedback and Profit Calculation. After the post-trading survey page, traders go to the feedback page, which shows the true state realized in the current period, and also reminds traders of their final asset and cash holdings. The page then

[2] Error messages are also triggered by trying to accept one's own offer and by trying to acquire more than 8 asset units.

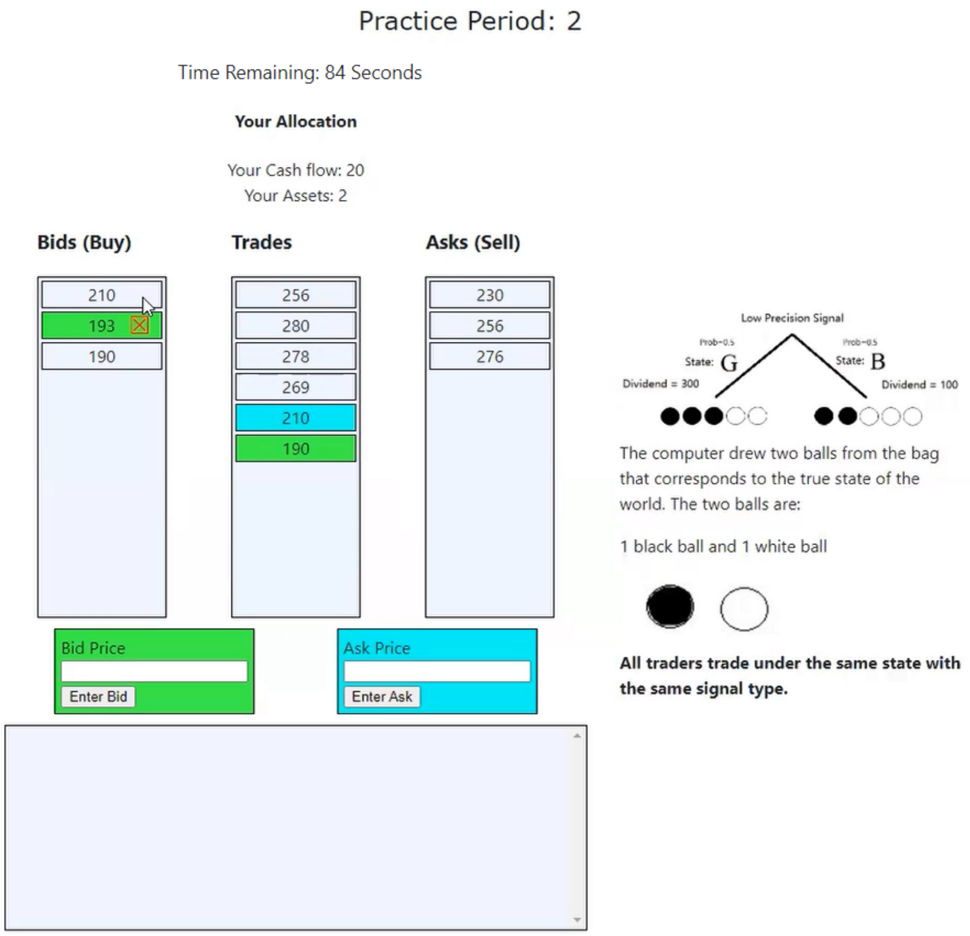


Figure 3.3: User Interface for Trading in Environment aL

computes trading profit as well as payments for responses to all the survey questions. As noted in the previous section, trading profit is $\pi = nd + R - C$. Payments for survey question are scaled to roughly equalize maximal payments across questions and so that their average over the $2 \cdot 3 = 6$ questions is about half of average trading profit. Let τ be a subject's average questionnaire payment in a given period. Then their total payoff for that period is $\pi + \tau$. Payoffs are summed over all 14 paid periods, and paid to the subject at the end of the session.[3]

[3]We adopt the common convention from market experiments of paying for all periods, and not just

3.3.2 Information Environments

Traders all face the same unknown state each period, but their private information is generated differently in different environments. In Environment aL, all traders receive independent low precision signals and are so informed, as in Figure 3.3. Environment aH is the same except that all signals are high precision, as in the right panel of Figure 3.1. Among the 14 periods in environment a sessions, half of the periods are randomly selected to be Environment aL, and the other half to be Environment aH.

In each of 14 periods in environment b, half the traders are randomly selected to receive high precision signals while the other half receive low precision signals; each trader gets H precision signals in 7 periods on average. The UI reminds them of that heterogeneity, and tells them their own signal type, H or L.

Environment c is exactly the same as b, except that the UI tells the trader nothing about the precision of other traders' private information other than that it may differ from their own. Traders' private signals are not labelled L or H, and therefore relative precision is unknown or ambiguous in Environment c.

We would expect traders tend to be more (less) overconfident if they know their own signal is more (less) precise than other traders', but such an effect may disappear when traders know nothing about the relative precision of their own signal.

3.3.3 Implementation Details

Twenty sessions were conducted online from May 2021 to June 2022. Using ORSEE (Greiner (2015)), 112 human subjects were recruited from our subject pool;

one period selected randomly. That convention harmonizes with our risk-neutral theoretical predictions.

they are predominantly undergraduates pursuing a variety of majors. Each session had 8 of these subjects facing a single information environment (a, b or c). Subjects in "inexperienced" sessions had no prior experience with any of these environments nor with analogous environments featured in the companion paper (Wang (2022)). Subjects with above average performance in such a session were invited to participate in a subsequent "experienced" session. Table 3.1 summarizes the entire set of sessions.

Environment	Number of Inexperienced Sessions	Number of Experienced Sessions
a	5	2
b	5	2
c	4	2

Table 3.1: Sessions Implementation

The software for running the sessions was built in the lab using the oTree platform (Chen et al. (2016)). The sum of earnings (in points) in the 14 paid trading periods is converted to US dollars at a pre-announced rate and paid to each participant, together with a \$7 guaranteed show-up fee (increased to \$10 in the last few sessions, as the available subject pool shrank due to Covid-19.) Average payments were approximately \$30 per subject for a two-hour session.

3.3.4 Hypotheses to Test

Those laboratory procedures allow us to recast our general research questions as the following set of testable hypotheses.

H1. Overconfidence is a consistent individual trait: the overconfidence indexes for each subject's pre-trade responses to Q1, Q2 and Q3 will be highly correlated.

This hypothesis deals with an important preliminary issue, but it is not well supported by previous literature. For example, Fan et al. (2021) finds a major gap between asset valuation (as in Q2) and elicitation of the underlying probabilities (as in Q1). We noted earlier that the standard literature distinguishes between judgemental overconfidence (as in Q1 and Q2) and self-enhancement (as in Q3). To test H1, we rely on the overconfidence index x for Q1 responses as defined in equation (3.2) and for Q2 responses as defined in equation 3.4. For Q3 we will set $x =$ actual rank - elicited expected rank.

H2. The post-trade deviations of traders' elicited beliefs and valuations from their fully-aggregated values P and V^* will be substantially smaller than their pre-trade deviations.

This hypothesis deals with another important preliminary issue, on whether market participation disseminates dispersed private information. Barron and Qu (2014) suggests the distribution of private signals plays an important role in affecting market efficiency. While traders in their market have perfect information about signal precision, traders in one of our market environments do not. We would expect dissemination to be better when more traders have high precision signals, and when relative signal precision is known.

H3. Informational efficiency (or information aggregation) will be enhanced by subject experience and known signal precision.

H3 addresses the last of our preliminary issues, and is based on natural conjectures. We will test it by regressing the metrics introduced in Section 3.2.1 (absolute deviation of asset price from V^*, and Page and Siemroth (2021)'s λ) on environment dummy variables.

H4. Overconfidence and underconfidence will both impair traders' overall profits.

This hypothesis reflects the Alchian (1950) tradition that markets punish deviations from full rationality. We shall test this hypothesis by regressing $|x|$, the absolute value of our overconfidence index, on traders' realized profits.

H5. (a) Takers on average will trade at less favorable prices than makers. Overconfidence will (b) enhance trading volume of both makers and takers but (c) impair their price favorability. Also, (d) more overconfident traders will be more likely to be takers and less likely to be makers.

This is a finer-grained version of the previous hypothesis that pertains to the CDA format for asset markets.

H6. Trading experience attenuates traders' overall overconfidence (measured by question 3) in all market environments.

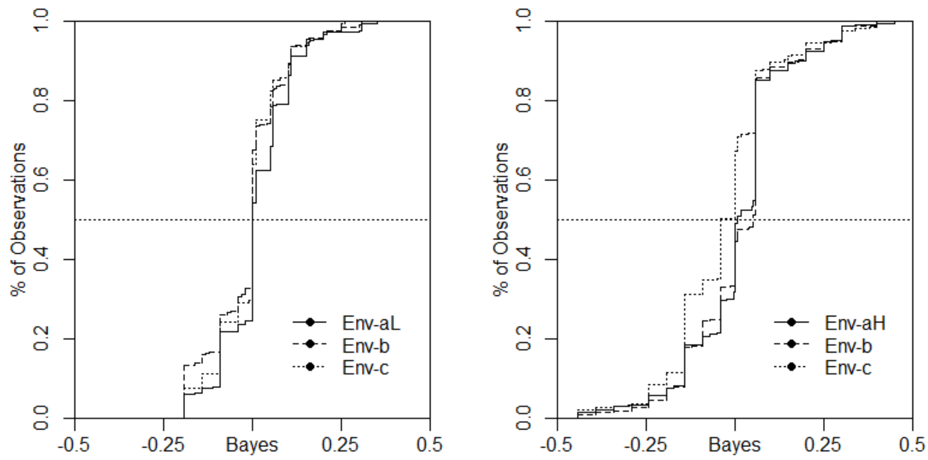
This pertains to our final research question, on the reciprocal effect of trading experience on overconfidence, and it reflects the other aspect of the Alchian (1950) tradition. We would like to use subjects' responses to all three questions but, as explained in previous section, it is infeasible to obtain post-trade overconfidence indexes for Q1 and Q2. Therefore, we shall test H6 only on the Q3 overconfidence index.

3.4 Results

We begin with descriptive statistics and tests of the preliminary hypotheses (H1-H3) in Sections 3.4.1 - 3.4.3. Section 3.4.4 examines how overconfidence impacts individual trader profits and profit components such as trading volume and the spread between buying and selling prices. Section 3.4.5 looks at the reciprocal impact of market experience on overconfidence.

3.4.1 Overconfidence Summary Statistics

Using pre-trade survey responses in sessions with inexperienced subjects, Figure 3.4 plots the cumulative distribution function (CDF) of the overconfidence index x for question 1. Neither panel shows much difference across market environments. It is reassuring to see that pre-trade beliefs seem largely unaffected by the market environment that subjects will face later in the period (and that they faced in earlier trading periods). This impression is reinforced by Wilcoxon Signed-Rank (WSR) test results presented in Table C2.



Cumulative distribution functions of overconfidence index x for Survey Question 1 prior to trading. Left panel is for subjects with low precision (L) signals, and right panel is for high precision (H) signals.

Figure 3.4: Q1 Pretrade Over-/Under-confidence.

The right panel shows that traders with high precision signal are, roughly speaking, as likely to be under- as over-confident. By contrast, the left panel shows rather little underconfidence among traders receiving a low precision signal. Again, that is reassuring, since there is less room for underconfidence when the Bayesian posterior P is closer to 0.5. Similar CDFs lead to similar conclusions for questions 2 and 3, and for experienced sessions, as can be seen in Appendix C.1.2.

Inexperienced Sessions					
	All Envs	Env-aL	Env-aH	Env-b	Env-c
Corr(Q1,Q2)	0.52***	0.42***	0.54***	0.55***	0.51***
Corr(Q2,Q3)	0.00	-0.08	0.09	-0.05	0.04
Corr(Q1,Q3)	0.05*	-0.04	0.11*	-0.01	0.10**
(Obs.)	(1427)	(247)	(256)	(509)	(415)
Experienced Sessions					
	All Envs	Env-aL	Env-aH	Env-b	Env-c
Corr(Q1,Q2)	0.36***	0.37***	0.43***	0.27***	0.46***
Corr(Q2,Q3)	0.09**	-0.11	0.15	0.06	0.21***
Corr(Q1,Q3)	0.07*	0.14	0.21**	-0.01	0.04
(Obs.)	(625)	(103)	(105)	(203)	(214)

Note: Corr(Qi,Qj) refers to the Pearson correlation between overconfidence index x for pre-trade question i responses and that for the same subject's pre-trade question j responses, $i, j = 1, 2, 3$. Numbers of observations are shown in parentheses for each environment and subject experience level. Asterisks ***,**,* respectively indicate p-values 0.01, 0.05, and 0.10.

Table 3.2: Overconfidence Correlations

Table 3.2 shows correlations between overconfidence measured in different survey questions. It supports H1 for Q1 and Q2, which is reassuring since one is a linear function of the other and both deal with judgemental overconfidence.[4] On the other hand, correlations are much lower with the third question, which deals with self-enhancement.

[4]This finding does not directly contradict Fan et al. (2021), since their finding of a gap between probability judgements and valuation judgements is based on overconfidence levels rather than overconfidence correlations. Appendices C.1.5 and C.1.6 show that our L-precision data exhibit a gap in overconfidence levels similar to theirs.

Result 3.4.1.1 *Pre-trade over- and under-confidence revealed in Q1 (beliefs about the state) is highly correlated with that revealed in Q2 (asset valuation), but neither is well correlated with that revealed in Q3 (rank in trading profit).*

We now turn to post-trade beliefs, and examine whether trading experience attenuates signed deviations $p-P$ of elicited probabilities p from fully-aggregated Bayesian posterior probabilities P. The black solid (and dashed) lines in Figure 3.5 show the CDFs of such deviations inexperienced sessions for pre- (and post)-trade elicitation p. For comparative purposes, the red dashed line shows deviations from P of individual traders' pre-trade true Bayesian posteriors (i.e., given only their personal pre-trade private signals). One gets the impression that trading experience has little impact in the low precision environment aL, but that it does indeed attenuate both positive and negative belief errors in the other environments. Appendix C.1.12 includes graphs for the other two questions and for experienced sessions, and they exhibit similar patterns.

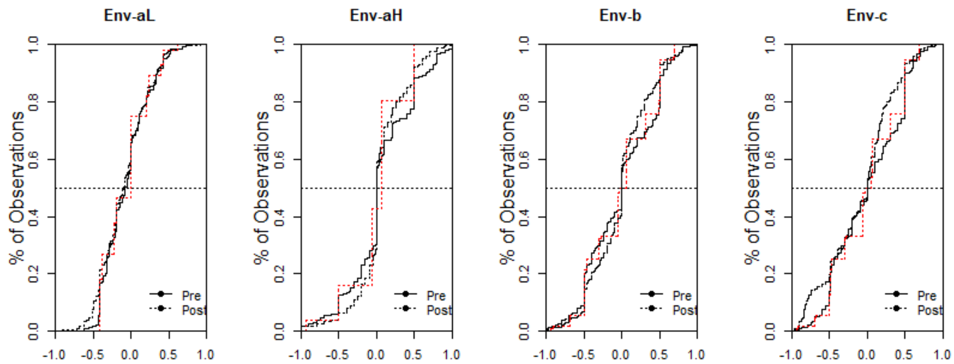


Figure 3.5: Deviations of elicited beliefs (Q1) from fully-aggregated Bayes posteriors.

Table 3.3 provides more formal support for these impressions. It reports p-values for the Wilcoxon signed rank test comparing absolute deviations of pre- and post-trade elicited beliefs from P. Except for env-aL, we can clearly reject the null hypothesis that post-trade deviations are no smaller than pre-trade deviations. Rejection is especially emphatic for Q1 and Q2 for experienced subjects in the mixed environments b and c.

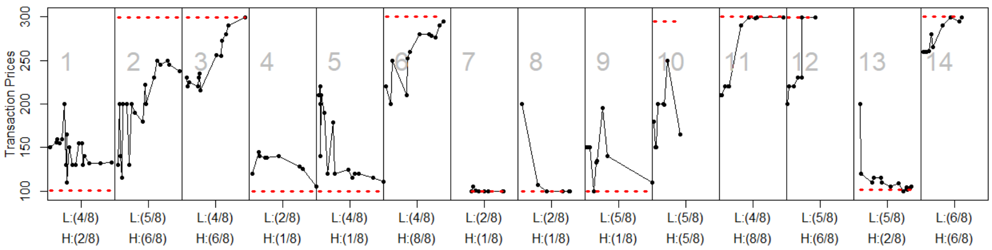
	Inexperienced Sessions				Experienced Sessions			
	Env-aL	Env-aH	Env-b	Env-c	Env-aL	Env-aH	Env-b	Env-c
Q1	0.1431	0.0020	0.0000	0.0112	0.2737	0.2733	0.0000	0.0000
Q2	0.7213	0.0000	0.0000	0.0045	0.4196	0.0367	0.0000	0.0000
Q3	0.3425	0.2608	0.0104	0.0387	0.1800	0.0227	0.0438	0.0000

Table 3.3: Wilcoxon Signed-Rank test p-values for Absolute Belief Deviations

Result 3.4.1.2 *Deviations from fully-aggregated Bayesian beliefs are significantly reduced after market participation when at least half the participants have high precision signals, especially in environments b and c with experienced subjects. On the other hand, market participation has little or no impact when all participants have low precision signals.*

3.4.2 Market Performance and Information Aggregation

Figure 3.6 illustrates price dynamics in our markets. In most of the periods for the particular session shown, transaction prices tend to move away from the prior expectation of 200 towards the fully-aggregated expected asset value (the red dashed horizontal line, usually near 300 or 100). Convergence failed in period 10, where the high precision signals were least informative (5 of 8 balls were black). Similar tendencies can be observed, but often to a lesser extent, in the other 19 sessions; see appendices C.1.7, C.1.8, C.1.9, and C.1.10.



Note: Connected black dots show sequences of transacted prices in each of the 14 periods of one session. Horizontal axis labels show aggregate private information, e.g., L(4/8) H(2/8) means that 4 of the 8 balls seen by Low precision traders were black, and 2 of 8 balls seen by high precision traders. Fully aggregated Bayes expected asset values are plotted as red dashed lines.

Figure 3.6: Transaction Prices in Env-b Inexperienced Session 2

To study convergence more formally, we adapt the methods of Noussair et al. (1995), and estimate equation (3.5) below. The dependent variable y is the pricing error $|\nu - V^*|$, the absolute value of the deviation of actual price ν from rational expectations price (or Bayesian posterior expected asset value given all realized private information) V^* in equation 3.3. The explanatory variables include session-dependent transitory terms $D_i(1/t)$,[5] which vanish asymptotically as the number of periods t gets large, and a session-independent permanent term $(t-1)/t$ that converges to 1.0. Those terms are interacted with treatment dummies ($\mathbb{1}_{aL}$, $\mathbb{1}_{aH}$, $\mathbb{1}_b$, and $\mathbb{1}_c$), with Environment c taken as the baseline.

$$\begin{aligned} y_{it} = &\sum_{i=1}^{5} \beta_{aLi} D_i \mathbb{1}_{aL}(\frac{1}{t}) + \sum_{i=1}^{5} \beta_{aHi} D_i \mathbb{1}_{aH}(\frac{1}{t}) + \sum_{i=1}^{5} \beta_{bi} D_i \mathbb{1}_{b}(\frac{1}{t}) + \sum_{i=1}^{4} \beta_{ci} D_i \mathbb{1}_{c}(\frac{1}{t}) + \\ &+ \beta_c(\frac{t-1}{t}) + \beta_{aL}\mathbb{1}_{aL}(\frac{t-1}{t}) + \beta_{aH}\mathbb{1}_{aH}(\frac{t-1}{t}) + \beta_b\mathbb{1}_b(\frac{t-1}{t}) + u \end{aligned} \tag{3.5}$$

Table 3.4 shows our estimates for the permanent terms in equation (3.5), which are intended to pick up what would happen when the empirical convergence process is complete. The "actual price" ν in the first column in the Table is the last transaction

[5]Here D_i is a session dummy variable. Equation 3.5 assumes 5 sessions for most treatments, as in the inexperienced data, but of course in the experienced data the sums are just $i = 1$ to 2.

	Inexperienced			Experienced		
VAR	P-Error (Last Price)	P-Error (Last Two Prices)	Spread	P-Error (Last Price)	P-Error (Last Two Prices)	Spread
β_c	86.07***	83.89***	27.93***	4.826	4.009	24.22**
	(17.74)	(17.04)	(6.377)	(17.17)	(17.67)	(10.39)
β_{aL}	-14.10	-7.945	2.451	69.73***	76.93***	33.81*
	(28.50)	(27.52)	(11.86)	(24.76)	(25.12)	(20.14)
β_{aH}	-50.33*	-38.46	-16.92	38.59	60.12**	15.79
	(28.51)	(28.03)	(12.07)	(27.14)	(29.95)	(18.54)
β_b	-5.693	2.435	-3.184	55.92*	56.65**	10.02
	(31.37)	(29.89)	(10.54)	(28.72)	(28.02)	(14.55)
Obs.	196	196	195	84	84	84
R^2	0.702	0.728	0.663	0.395	0.424	0.620

Robust standard errors in parentheses; *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 3.4: Regressions of Convergence

price each period, while in the second column ν is the average of the last two transaction prices. The third column replaces the dependent variable pricing error by "Spread," the difference between the last best ask and the last best bid.

The results may seem a bit surprising. In inexperienced sessions, the baseline environment c is not at all conducive to convergence to the fully aggregated asset value (typically near 100 or 300); the estimated price error is over 80 and highly significant, and the estimated spread of almost 28 is also quite large and significant. Price errors on average are considerably less (by around 40 or 50, i.e., about half as large) in the high precision treatment aH, but there is so much variability across sessions that the reduction is at best marginally significant. The other environments seem to have somewhat better convergence than baseline but not significantly so. By contrast, convergence in experienced sessions is remarkably good for baseline environment c: price errors are typically only about 5. Spread in env-c decreases only slightly, however, relative to inexperienced sessions. Not surprisingly, price errors are considerably larger in the other experienced treatments, and the difference is highly significant for aL, the environment with least precise private information.

To check the robustness of these results, we examine alternative measures of information aggregation. Recall the discussion in Section 3.2.1 of Page and Siemroth (2021)'s λ. The same paper also defines "price accuracy" as $\psi = min\{1, (p - 0.5)/(P - 0.5)\}$, where (only in the current paragraph) p and P are the actual price ν and fully aggregated value V^* rescaled from their natural range [100, 300] to the unit interval [0,1]. Thus this notion of price accuracy is in some ways reminiscent of our overconfidence index x, but the minimum operator ensures that ψ cannot exceed 1, thereby treating prices that overreact to information as if they reacted correctly. To estimate average accuracy ψ of the (average of the last two rescaled) observed prices p_m in reaching fully aggregated rescaled prices P_m in period $m = 1, ..., M$, one reports the fitted coefficient in the regression

$$Y_m = \psi(P_m - 0.5) + \epsilon_m \tag{3.6}$$

where

$$Y_m = \begin{cases} min\{p_m - 0.5, P_m - 0.5\}, \text{ if } P_m > 0.5, \\ max\{p_m - 0.5, P_m - 0.5\}, \text{ if } P_m < 0.5. \end{cases}$$

Table 3.5 collects the results. Consistent with the previous table, it indicates that markets aggregate information best in experienced sessions in environment c, and also quite well in aH, even in inexperienced sessions. The λ measure indicates that experienced sessions do better than inexperienced sessions in all environments, and agrees that aggregation is worst in inexperienced aL sessions.

Data	Inexperienced λ	Experienced λ	Inexperienced ψ	Experienced ψ
Env_aL	0.06	0.25	0.05	0.24
	[0.00,0.31]	[0.00,0.50]	[-0.16,0.22]	[0.03,0.59]
Env_aH	0.12	0.25	0.69	0.64
	[0.12,0.31]	[0.12,0.37]	[0.59,0.78]	[0.34,0.88]
Env_b	0.12	0.12	0.47	0.5
	[0.12,0.25]	[0.12,0.37]	[0.34,0.57]	[0.30,0.69]
Env_c	0.12	0.25	0.35	0.72
	[0.00,0.12]	[0.18,0.50]	[0.26,0.43]	[0.52,0.88]

Note: The 95% confidence intervals below the estimated λ and ψ are calculated via the non-parametric percentile bootstrap method, which resamples (with replacement) the market periods within each environment 100 times to determine the final confidence intervals. We also estimate λ and ψ using the last transaction price in each period, the results are similar, please see table C1.

Table 3.5: Estimated λ and ψ (Last Two Prices)

Combining the lessons of Tables 3.4 and 3.5, we have the following

Result 3.4.2.1 *Informational efficiency (or information aggregation) is generally enhanced by subject experience, especially in the ambiguous mixed environment c. It is impaired when all signals are low precision, as in environment aL.*

3.4.3 Impact of Overconfidence on Prices

To assess the impact of average overconfidence $\bar{x}$ in traders' pre-trade responses to Q1 on actual asset price ν (last or average of last two) in any given period, we form the dependent variable $\mathcal{R} = (\nu - V^*)sgn(V^* - 200)$ and run the regression

$$\mathcal{R} = \beta_0 + \beta_c\bar{x} + \beta_{aL}\bar{x}\mathbb{1}_{aL} + \beta_{aH}\bar{x}\mathbb{1}_{aH} + \beta_b\bar{x}\mathbb{1}_b. \tag{3.7}$$

Actual price over- (under-) reacts relative to fully aggregated price V^* to the extent that $\mathcal{R}$ is positive (negative), and traders in a given period are on average over- (under-

) confident to the extent that $\bar{x}$ is positive (negative). The symbols of $\mathbb{1}_{aL}$, $\mathbb{1}_{aH}$, $\mathbb{1}_{b}$ are dummy variables for treatments other than environment c, which is the baseline.

	Inexperienced				Experienced			
	Last One		Last Two		Last One		Last Two	
VARIABLES	Q1	se	Q1	se	Q1	se	Q1	se
β_c	-100.9	(155.0)	-75.77	(148.3)	-292.2*	(149.4)	-344.3**	(158.8)
β_{aL}	380.2	(305.2)	301.6	(279.2)	409.7	(462.0)	304.2	(449.8)
β_{aH}	50.14	(206.9)	34.09	(195.1)	-88.58	(199.5)	38.11	(231.6)
β_b	457.6**	(215.2)	376.4*	(211.3)	957.7**	(410.0)	1,035***	(370.6)
Constant	-52.19***	(16.05)	-53.48***	(15.14)	-30.65**	(13.89)	-40.67***	(15.29)
Observations	196		196		84		84	
R-squared	0.259		0.262		0.121		0.121	

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 3.6: Coefficient Estimates for Equation (3.7)

Table 3.6 reports the results. The significantly negative constant coefficients suggest that we generally see under-reaction. The average level $\bar{x}$ of overconfidence across traders is typically small, and it has insignificant impact in the inexperienced sessions except in the unambiguous mixed environment b. In that environment, overconfidence tends to partially offset under-reaction, i.e., it pushes prices towards V^*. We see the same directional impact in experienced sessions with environment b. By contrast (but consistent with Hypothesis H3) in experienced sessions with the ambiguous environment c, overconfidence significantly (at $p < 0.05$) reduces overreaction (or increases underreaction). Appendix C.1.14 reports similar results for overconfidence in Q2 responses, albeit with lower significance levels. It also reports generally insignificant results for Q3 overconfidence.

Result 3.4.3.1 *Asset prices generally underreact to aggregate information. In experienced sessions, trader overconfidence offsets that effect in mixed precision environment* *b, but not in ambiguous environment c.*

General underreaction is not surprising since the target V^* is so often very close to an endpoint, $V = 100$ or 300. The contrast between the unambiguous (b) and ambiguous (c) mixed precision environments may arise from low-precision traders' greater reliance on observed market prices in the unambiguous environment. We see that overconfidence in the experienced ambiguous environment pushes towards underreaction, consistent with Hypothesis H3.

3.4.4 Overconfidence and Individual Trading Performance

We now turn to the question of how overconfidence affects individual traders' market behavior. Here the most relevant elicitation is Q2, on subjective asset valuation, since it pertains most directly to trading strategies. Thus the overconfidence index $x = (v - V^*)sgn(V^* - 200)$ here is from equation (3.4). To estimate the impact on individual overall trading profit $\pi = nd + R - C$, we break the sample into overconfident ($x > 0$) vs underconfident ($x < 0$) trader-periods. To get consistent signs, we use the absolute value of x in the following regression.

$$\pi = \beta_0 + \beta_c|x| + \beta_{aL}|x|\mathbb{1}_{aL} + \beta_{aH}|x|\mathbb{1}_{aH} + \beta_b|x|\mathbb{1}_b \tag{3.8}$$

Table 3.7 reports the coefficient estimates (and robust standard errors). In the ambiguous precision baseline treatment c, being overconfident tends to impair profits, quite significantly so in experienced sessions. In that environment, being underconfident boosts profits significantly in inexperienced sessions, but the effect size shrinks consider-

VARIABLES	Inexperienced Over	se	under	se	Experienced Over	se	under	se
β_c	-0.828	(1.053)	2.779***	(0.857)	-4.318***	(1.213)	0.772	(1.295)
β_{aL}	0.726	(1.278)	-0.329	(1.954)	4.015***	(1.492)	-1.096	(2.661)
β_{aH}	-1.691	(1.246)	-3.870***	(1.146)	0.872	(1.949)	-0.202	(2.072)
β_b	0.745	(1.240)	-0.156	(1.196)	4.983***	(1.482)	3.712**	(1.786)
Constant	472.8***	(45.91)	401.8***	(59.40)	445.6***	(43.56)	443.7***	(104.6)
Observations	714		523		310		195	
R-squared	0.020		0.065		0.084		0.076	

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: Coefficient estimates for the restricted sample of overconfident (resp underconfident) traders are reported in the columns labeled "Over" (resp. "Under").

Table 3.7: Coefficient Estimates for Equation (3.8)

ably and significance disappears in experienced sessions. In homogeneous environment a, overconfidence boosts profits in low precision experienced periods, while underconfidence impairs profits in high precision inexperienced periods. In the unambiguous heterogeneous precision environment b, both under and over-confidence seem to boost experienced trader profits.

Result 3.4.4.1 *Overconfidence significantly impairs (resp. improves) experienced traders' profits in environment c (resp. b and aL) but elsewhere has little impact on overall trader profits. Underconfidence may boost experienced trader profits in environment b and inexperienced trader profits in environment c, but elsewhere has insignificant or negative impact.*

These apparently diverse findings mostly seem to jibe with the previous result that prices under-react more in environment c than in b. Perhaps being overconfident is more profitable when prices are less inclined to underreact.

For a finer-grained analysis, we now consider key observable elements of a trader's performance: their trading volume, their tendency to buy low and sell high, and their tendency to act as a market maker or taker. A contingency table sets the stage

	# of Sell: 0	# of Sell: 1	# of Sell: 2	# of Sell: More	Sum of Obs.
# of Buy: 0	79	112	300	0	491
# of Buy: 1	99	105	128	110	442
# of Buy: 2	77	57	42	80	256
# of Buy: More	129	43	48	159	379
Sum of Obs.	384	317	518	349	1568

Note: Entries report the overall number of instances (in all environments) in which a trader buys and sells specified numbers of units in a single period.

Table 3.8: Instances of Buy and Sell Volume

for that analysis. Table 3.8 shows that fairly often (in 300 of 1568 instances) a trader never buys and simply sells both her endowed shares. In about 10% of all instances the trader is an active market maker, selling at least 2 shares and buying at least 2 shares, and in about 5% of instances the trader does not trade at all. Appendix C.1.18 shows that the same general patterns are also seen in most environments separately.

We shall now examine more carefully the 772 instances where a trader both buys and sells in a single period. For those data, we run regression 3.9 below to study whether overconfidence coincides with trade at less favorable prices. The dependent variable is G, the "Sell-Buy-Gap" defined as the average price at which a trader sold asset units in a given period minus the average purchase price, i.e., their success in "buying low and selling high." Explanatory variables are the pre-trade Q1 overconfidence index x together with the usual treatment dummies.

$$G = \beta_0 + \beta_c x + \beta_{aL} x \mathbb{1}_{aL} + \beta_{aH} x \mathbb{1}_{aH} + \beta_b x \mathbb{1}_b \tag{3.9}$$

Table 3.9 reports the regression results separately for makers (instances in which the trader had both an accepted bid and an accepted ask) and takers (where a trader accepted both a posted bid and a posted ask). The constant terms indicate that the effective bid/ask spread is about 20 in inexperienced sessions and is considerably wider (!) in experienced sessions. Overconfidence seems to help makers and hurt takers slightly in baseline environment c, but less so in the other environments. In low information environment aL, the effect is reversed, significantly so for inexperienced makers.[6] Parallel (but weaker) results are reported in Appendix C.1.17 for x computed using Q2 responses.

	Inexperienced				Experienced			
VARIABLES	Maker	se	Taker	se	Maker	se	Taker	se
β_c	55.09*	(29.05)	-11.39	(32.64)	9.739	(42.91)	-33.16	(51.16)
β_{aL}	-140.1***	(41.55)	-30.50	(53.56)	-63.25	(79.74)	-48.54	(190.6)
β_{aH}	-72.08*	(40.98)	-12.65	(59.58)	-94.02	(64.64)	105.6	(97.77)
β_b	-23.50	(36.16)	2.887	(37.67)	-44.61	(62.75)	-17.94	(64.46)
Constant	19.33***	(4.898)	-20.33***	(5.915)	44.28***	(13.74)	-78.87***	(24.81)
Observations	374		289		149		107	
R-squared	0.076		0.060		0.257		0.137	

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 3.9: Overconfidence on Sell-Buy-Gap

Result 3.4.4.2 *Takers tend to trade at less favorable prices than do makers. Being more overconfident may improve (resp. worsen) makers' (resp. takers') transaction prices in ambiguous environment c, but the effect may be partially or fully reversed in most other environments, where traders know the precision of other traders.*

It is comforting to see that makers trade at favorable prices, since their viability

[6] Readers might wonder why the experienced maker and taker constant coefficients don't approximately offset each other. The discrepancy seems largely due to a single outlier, a trader with $x = -34$ and a sell-buy-gap = -159.

depends on it. Of course, less favorable prices for takers is the other side of the same coin. The result suggests that overconfidence may help makers better control transaction prices when their customers are unsure of their relative precision.

Of course, trading volume is the other component of profitability. To examine it, we run similar regressions with dependent variable M = the number of a maker's offers accepted in a given period:

$$M = \beta_0 + \beta_c \cdot x + \beta_{aL} \cdot x \cdot \mathbb{1}_{aL} + \beta_{aH} \cdot x \cdot \mathbb{1}_{aH} + \beta_b \cdot x \cdot \mathbb{1}_b. \tag{3.10}$$

Table 3.10 reports the results. Appendix C.1.16 reports parallel results on taker volume. In general, the patterns of taker volume are not systematic and less significant.

Result 3.4.4.3 *In experienced sessions, overconfidence reduces maker volume in mixed precision environments b and c, but increases it in the homogeneous environments aL and aH. Patterns are generally not significant in inexperienced sessions.*

An interpretation is that the impact of overconfidence on the buy-sell gap noted in the previous result is offset by its impact on trading volume, resulting in a mixed overall impact on trader profits.

3.4.5 Impact of Market Experience on Overconfidence

Finally, we examine the reciprocal effect of market experience on traders' overconfidence. Recall that it is infeasible to properly measure post-trade overconfidence for Q1 and Q2. Therefore Figure 3.7 below, and the corresponding figure for experi-

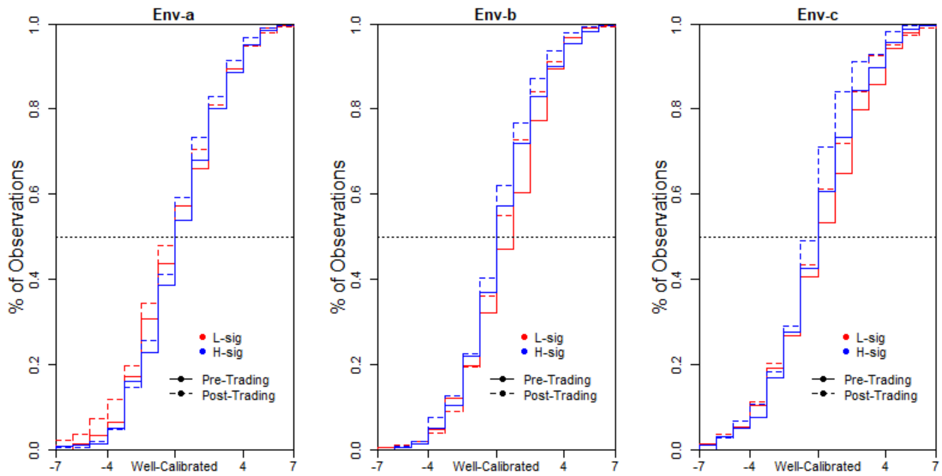
VARIABLES	Inexperienced Q1	Inexperienced Q2	Inexperienced Q3	Experienced Q1	Experienced Q2	Experienced Q3
β_c	0.0909	0.00290	-0.00599	-2.342**	-0.00478	-0.0997**
	(0.698)	(0.00233)	(0.0297)	(0.973)	(0.00325)	(0.0438)
β_{aL}	-0.319	-0.00921***	-0.0588	4.198**	0.00667	0.0644
	(1.136)	(0.00350)	(0.0491)	(1.776)	(0.00497)	(0.0668)
β_{aH}	0.557	0.000404	-0.0981**	3.730**	0.00502	0.0805
	(1.116)	(0.00325)	(0.0484)	(1.486)	(0.00548)	(0.0745)
β_b	-0.440	-0.00428	-0.0250	0.144	0.00594	0.0954
	(0.965)	(0.00330)	(0.0442)	(1.540)	(0.00442)	(0.0646)
Constant	1.688***	1.761***	1.708***	0.969***	0.904***	0.942***
	(0.136)	(0.140)	(0.132)	(0.0968)	(0.109)	(0.0944)
Observations	1,526	1,464	1,568	663	631	672
R-squared	0.094	0.104	0.099	0.107	0.097	0.099

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Note: We run the regression for different survey questions. "Q1", "Q2", and "Q3" means x is measured by question 1, question 2, and question 3, respectively.

Table 3.10: Overconfidence on Maker Volume

enced sessions in Appendix C.1.4, use only Q3 responses. These figures suggest, and



Note: More negative (resp. positive) entries in each panel indicate more under-(over-)confident traders.

Figure 3.7: Pre vs Post-Trading Overconfidence (Question 3)

the Wilcoxon summed-rank tests reported in Table C4 confirm, that the distribution of overconfidence is largely unaffected by market participation.

Result 3.4.5.1 *Trading experience does not substantially attenuate traders' overconfidence (measured by question 3) in any of the environments we examined.*

This finding is consistent with some previous literature, e.g., Hoffman and Burks (2017), Huffman et al. (2019).

3.5 Conclusion

Our results can be summarized briefly. First, the preliminary question of whether overconfidence is a coherent personal trait, we have mixed results. Correlations indicate no substantial gap between judgemental overconfidence in valuation (Q2) and in assessing probabilities (Q1). On the other hand, consistent with the standard distinction between judgemental and self-enhancing overconfidence, we find a substantial gap between our self-enhancement measure (Q3) and either sort of judgemental overconfidence.

We add a bit of nuance to the vast set of previous research on information dissemination and aggregation in asset markets. When at least half our traders received high precision signals prior to trade, their post-trade beliefs better reflect aggregate information, despite the fact that the aggregate information is more precise in such cases. On the other hand, market participation has little or no impact on beliefs when only low precision pre-trade private information was available. Likewise for information aggregation: asset prices better reflect aggregate private information when at least half of it is high precision (especially in the ambiguous precision environment c), but not when it is all low precision.

Our main results concern the impact of overconfidence on market performance and the reciprocal impact of market experience on overconfidence. We find that trader overconfidence can sometimes improve price efficiency (i.e., information aggregation). This improvement arises mainly from mitigating under-reaction of asset price to aggregate information. Such improvement occurs in our unambiguously mixed precision environment, but not in the most challenging environment with ambiguous mixed precision.

Other aspects of market performance, such as trader profitability and trading volume, are also impacted by trader overconfidence. For example, overconfident traders earn lower profits in the ambiguous precision environment when their peers are experienced. In those same environments, market makers may be able to increase the effective spread while reducing trading volume.

A bit to our surprise, we are unable to detect any systematic impact of market exposure on traders' overconfidence. An interpretation is that the first part of the Alchian (1950) story is less reliable than one might think. Traders are not always punished for irrational over (or under) confidence. Given that, it is perhaps less surprising that market experience fails to eliminate over (or under) confidence.

Our experiment is exploratory, but we hope that it will encourage future research on interactions between trader overconfidence and asset market performance. Such work will help sharpen theoretical understanding, and may also help financial market regulators who seek to improve access and market performance.

Appendix A

Chapter 1 Appendix

A.1 Figures

A.1.1 User Interface

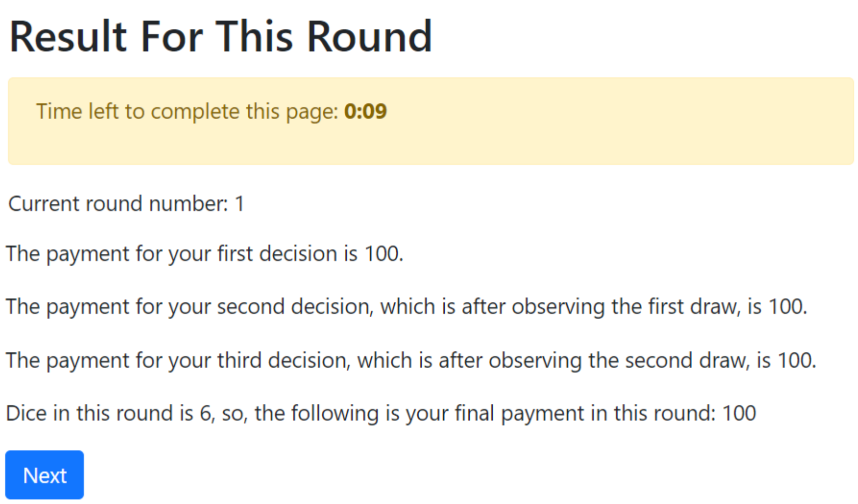


Figure A1: UI for Payment In Each Round

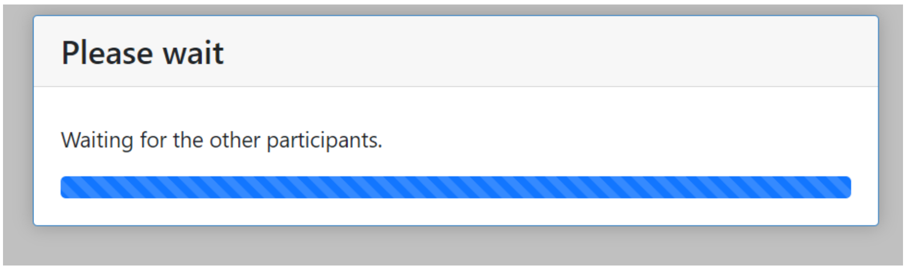


Figure A2: UI for Waiting Page

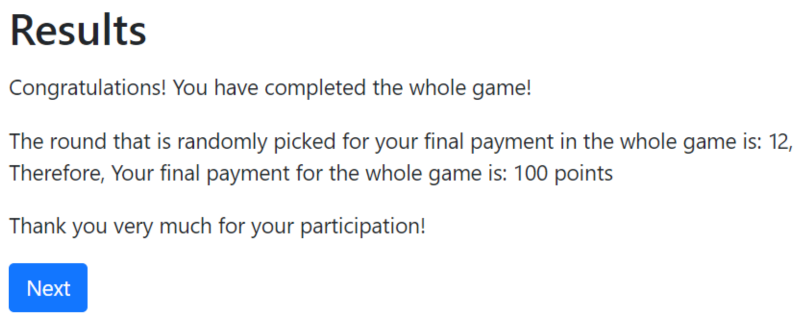
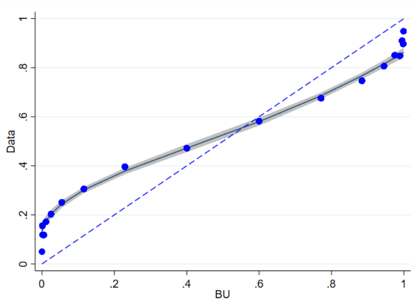
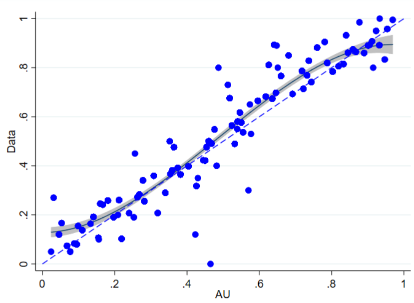


Figure A3: UI for Results

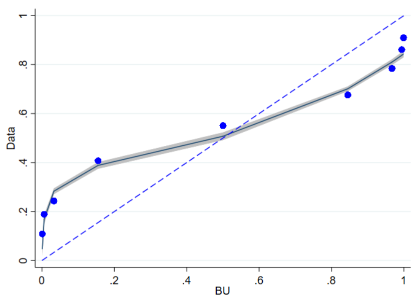
A.1.2 Goodness of Fit (Second Order Polynomial with 95% Confidence Interval)



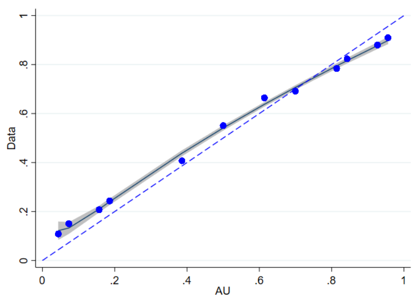
A: Bayesian vs Data (Parm-A)



B: Average vs Data (Parm-A)



C: Bayesian vs Data (Parm-B)



D: Average vs Data (Parm-B)

Figure A4: Goodness of Fit: Bayesian vs Average

A.1.3 Average Model vs LR(Ave) Model (Parm-B)

One thing interesting to notice for Parm-B sessions is approximately one third of the subjects' beliefs could be fitted well by Average model and LR(Ave) model, such as subjects 3, 8, 10-12, 14, 29, 31, 33, 36, 38, 39, 41, 43. Average model and LR(Ave) model did similar job in Parm-B is because their predictions are similar to each other as explained, also see the following figure A5.

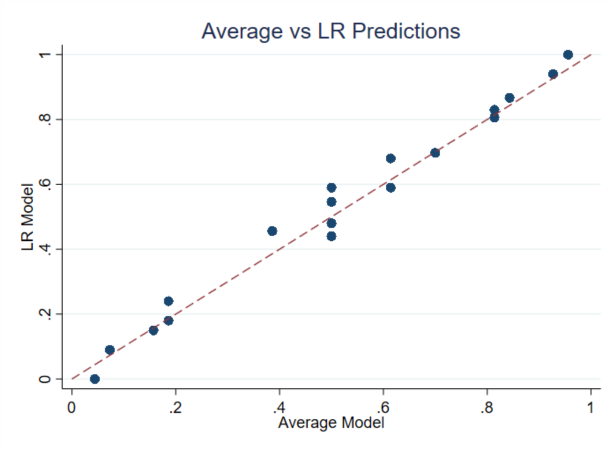


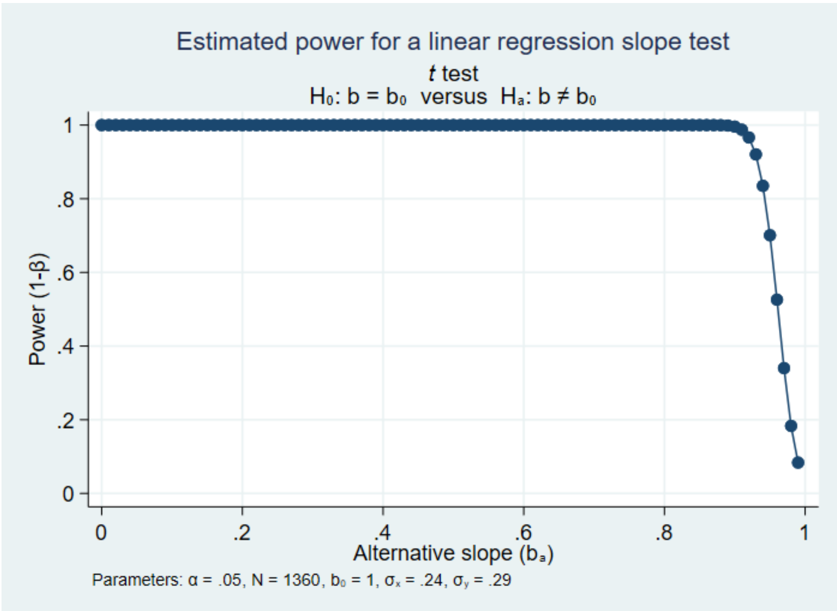
Figure A5: Average Model vs LR Model in Parm-B

A.1.3.1 Statistical Power for β_1 in Parm-A Sessions Regression

For Parm-A sessions, the hypothesis tests of Average model's prediction on elicited data tell us that we fail to reject that the fitted line has an intercept of 0 and slope of 1. One might think there could be a type two error, which means I might fail to reject the null while the alternative is true. To alleviate this doubt, I calculated the statistical power for β_1 on Average's regression:

$$Power = 1 - prob(\text{Fail to Reject} H_0 | H_1 \text{is true}) = prob(\text{Reject} H_0 | H_1 \text{is true})$$

The power varies with the alternative β_1. The plot for power against different alternative β_1 is provided in figure A6. Given the current data size, the power remains at a high level for most of the alternative β_1s. For example, if we have alternative that $\beta_1 = 0.93$, we will have a power higher than 90%. This indicates that the probability of making type two error is low.



Note: Conditional on the current data size, the power for β_1s stayed at a high level for most of the alternatives.

Figure A6: Statistical Power for β_1

A.2 Tables

A.2.1 Tables for Regressions

In table A1, I provide the same regression as table 2 but with outliers dropped. The definition of the outliers is: one observation is an outlier if its elicited probability is 50% higher/lower than either of the model's predictions. This means, for example, if any of the prediction predicts a posterior close to 0%(100%), the subject reported at least a 50% higher(lower) number into the computer, which is an obvious outlier. Dropping those outliers causes me to lose about 1.9% of the observations for Parm-A sessions and lose about 8.2% for Parm-B sessions. Table A1 says that Bayesian's prediction does not perform very well since it significantly deviates from the 45-degree line. Average model performs better in both Parm-A and Parm-B sessions in this case, from the hypothesis test results, we could see that, after dropping the outliers, we fail to reject the fitted line for Average model is 45-degree line in both parm-A and parm-B sessions. The Average model's better performance in Parm-B sessions after dropping the outliers also supported the previous explanations of higher standard deviation in Parm-B sessions.

A.2.2 Hypothesis Tests on Individuals' Signals

Explanation:

Table A2 and A3 show us the mean of observed black balls from the first and the second draw in Parm-A sessions. The second column in table A2, for example, shows, in case 1, when people exhibit Bayesian, the mean of the number of black balls each subject observes. Similarly, the third column shows, in case 1, when people exhibit Average,

Sessions VARIABLES	(1) Parm-A Data	(2) Parm-A Data	(3) Parm-A Data	(4) Parm-A Data	(5) Parm-B Data	(6) Parm-B Data	(7) Parm-B Data	(8) Parm-B Data
Bayesian	0.663***				0.649***			
	(0.0504)				(0.0471)			
Average		1.005***				0.922***		
		(0.0708)				(0.0647)		
LR(Ave)			1.391***				0.920***	
			(0.0925)				(0.0654)	
LR(Bayes)				0.837***				0.663***
				(0.0586)				(0.0474)
Constant	0.118**	-0.0512	-0.265**	0.00842	0.273**	0.140*	0.129*	0.257**
	(0.0249)	(0.0350)	(0.0472)	(0.0306)	(0.0289)	(0.0373)	(0.0386)	(0.0297)
Observations	1,334	1,334	1,334	1,334	1,579	1,579	1,579	1,579
R-squared	0.766	0.767	0.762	0.775	0.813	0.820	0.820	0.823
RMSD	0.1424	0.1420	0.1435	0.1395	0.1308	0.1284	0.1284	0.1273
p for $\beta_0 = 0$	0.0089	0.2167	0.0049	0.7969	0.0007	0.0198	0.0289	0.0010
p for $\beta_1 = 1$	0.0026	0.9429	0.0134	0.0501	0.0017	0.2939	0.2875	0.0021
p for joint test	0.0026	0.9429	0.0134	0.0501	0.0017	0.2939	0.2875	0.0021

Robust standard errors in parentheses
*** p<0.001,** p<0.01,* p<0.05

Table A1: Regressions Dropping the Outliers

First Draw (Parm-A)				
	Case 1	Case 1	Case 4	Case 4
	Bayesian	Average	Bayesian	Average
id_1	10	9.6	5.67	5.8
id_2	11	10	5.33	5.25
id_3	8.6	8.5	6.25	4.62
...	...	...	...	...

Note: The number in this table is the mean of number of black balls subjects observe in each case when they exhibit Bayesian or Average.

Table A2: Mean of Number of Black Balls Observed by Subjects (First Draw Parm-A)

Second Draw (Parm-A)				
	Case 1	Case 1	Case 4	Case 4
	Bayesian	Average	Bayesian	Average
id_1	6	6.4	3.17	2.8
id_2	9	6.67	2.33	3.5
id_3	6.2	7.25	3.5	2.88
...	...	...	...	...

Note: The number in this table is the mean of number of black balls subjects observe in each case when they exhibit Bayesian or Average.

Table A3: Mean of Number of Black Balls Observed by Subjects (Second Draw Parm-A)

the mean of the number of black balls each subject observes. Therefore, we have a distribution of the mean of black balls subjects observe when they exhibit Bayesian and Average in case 1. Then we can do a two-sample t-test to see if the number of black balls subjects observe significantly different from each other when they exhibit different behaves, and thus we could know if the reason that subjects behave like Bayesian and Average is from the balls they observe are different. I am omitting examples from Parm-B sessions since table A2 and A3 aim to have the reader understand the t-test, the important t-test results are provided below in table A4 and A5. I denote the mean of the number of black balls subjects observe when exhibiting Bayesian/Average as $\mu_{Bayesian}/\mu_{Average}$.

First Draw		Second Draw	
Case 1	Case 4	Case 1	Case 4
H0:$\mu_{Bayesian}$=$\mu_{Average}$ H1:H0 not true	H0:$\mu_{Bayesian}$=$\mu_{Average}$ H1:H0 not true	H0:$\mu_{Bayesian}$=$\mu_{Average}$ H1:H0 not true	H0:$\mu_{Bayesian}$=$\mu_{Average}$ H1:H0 not true
P-value= 0.5007	P-value=0.03601	P-value= 0.8605	P-value= 0.4393
Fail to Reject H0	Reject H0 at 95% level	Fail to Reject H0	Fail to Reject H0

Table A4: Two-sample T-test (Parm-A)

Explanation:

First Draw		Second Draw	
Case 1	Case 4	Case 1	Case 4
H0:$\mu_{Bayesian}$=$\mu_{Average}$ H1:H0 not true	H0:$\mu_{Bayesian}$=$\mu_{Average}$ H1:H0 not true	H0:$\mu_{Bayesian}$=$\mu_{Average}$ H1:H0 not true	H0:$\mu_{Bayesian}$=$\mu_{Average}$ H1:H0 not true
P-value= 0.1298	P-value=0.5584	P-value= 0.00565	P-value= 0.000
Fail to Reject H0	Fail to Reject H0	Reject H0	Reject H0

Table A5: Two-sample T-test (Parm-B)

From the t-test result in table A4, we could see that only in case 4's first draw did subjects observe a different pattern. In other cases, subjects observe a similar number of black balls and white balls in each signal, so the reason they exhibit different behaviors does not come from the pattern in the signals they observe in Parm-A sessions. The only case that we reject the null hypothesis is case 4 in the first draw. Although we reject our null in this case, it is a little counter-intuitive. Because in case 4, which means p_1 and p_2 both smaller than 0.5, we would expect the less black balls indicate it's more likely that the lottery worth 0, and thus subjects should be more confident about the true state of the world and thus behave more like a Bayesian. However, the data shows subjects behave like Bayesian when they observe more black balls. One reason could be, in case 4, subjects realized it's more likely that the lottery worth 0, so, to prevent themselves from losing money, they might enter low probability even though they saw many black balls in the signal so that they are able to exchange their initial lottery with the N-lottery, which could have a higher probability to worth 100. Table A5 shows us the results of Parm-B sessions. As we could see, we reject the hypothesis that $\mu_{Bayesian} = \mu_{Average}$ only for the first draw. $\mu_{Bayesian}$ and $\mu_{Average}$ are significantly different from each other in the second draw. The potential explanation for this could be in Parm-B sessions, since the number of balls drawn reduces and balls in the bag are more informative, so subjects might feel "enough informed" from the first signal, and thus put less weight on the second signal. This could also be the potential explanation for the pattern in figure 8, where Parm-B sessions' graph has a second peak at around 0.4. When subjects feel less informed from the second draw and thus put less weight (or even ignore) on the information from it, the three behaviors (Bayesian, Average, and No-response) could emerge even when the absolute difference of black balls ratio between two signals are high, which is, in this case, the ratio of 0.4.

A.3 Math Proofs

A.3.1 Shannon's Entropy and Parm-B Sessions

Since subjects' task is to guess the probability that their lottery worth 100, so the entropy I need to calculate should be conditional entropy, which is the probability that the lottery worth 100 conditional on the pair of draws subjects observe and I denote this as:

$p(y|x_i)$, where y is the event that lottery worth 100 (v=100), so y^c is the event that the lottery worth 0 (v=0). x_i is the pair of draws the subject observed. For example, in the Parm-A sessions, we have 15 draws in the first signal, and 10 draws in the second signal. So, all possible pairs of draws that subjects might observe are in the following table, and the total possible draws are $16 \times 11 = 176$. $p(x_i)$ denotes the probability that pair x_i is drawn.

	Signal 1		Signal 2	
	# black	# white	# black	# white
x_1:	0	15	0	10
x_2:	0	15	1	9
x_3:	0	15	2	8
...	...	...	...	
x_i:	1	14	0	10
...	...	...	...	
x_{176}:	15	0	10	0

Conditional Entropy in this case should be:

$$
\begin{aligned}
H(Y|X) &= -\sum_{i=1}^{176} p(x_i)H(Y|X = x_i) \\
&= -\sum_{i=1}^{176} p(x_i)\Bigg(\sum_{Y\in\{y,y^c\}} p(Y|x_i)log_2 p(Y|x_i)\Bigg) \\
&= 0.5065342
\end{aligned}
$$

For Parm-B sessions, I set the black/white ratio in the good bag equals to 7:3, and the black/white ratio in the bad bag equals 3:7. Then I need to decide the number of draws for each signal. The lowest number of draws for the first signal is 3 and the highest is 15. The lowest number of draws for the second signal is 2 and the highest is 10. Therefore, there will be $(15 - 3 + 1) \times (10 - 2 + 1) = 117$ different protocols. Then I calculated conditional entropy for all the different protocols and then found the 10 protocols that have the closest entropy value as the Parm-A session's. They are:

First Draw	Second Draw	Entropy
4	2	0.4954
3	3	0.4954
3	2	0.5535
5	2	0.4440
3	4	0.4440
4	3	0.4440
4	4	0.3985
6	2	0.3985
3	5	0.3985
5	3	0.3985

Since I need to have the number of draws different for each signal, the protocols with the same number of draws in each signal are not good options. Also, since the number of draws is just a few this time, even number of draws will often result in 1/1 and or 2/2 outcomes, which leave Bayesian posteriors unchanged. So, based on the above concerns, the first signal draw 5 balls and the second signal draw 3 balls is chosen for the Parm-B sessions.

A.3.2 Proof of Expected Payoff

The probability (P_N) that the N-lottery worth 100 is a random number from 0 to 99. For baseline, since subjects haven't received any signal, they will report 50% as their belief. In this case, there will be 50% chance that $P_N < 50\%$ and thus the subjects will hold their initial lottery, which has expected payoff of $100 \times \frac{1}{2}$. Also, there is 1% chance that $P_N = 51 > 50\%$ and thus subjects will hold the N-lottery, whose expected payoff is $100 \times 51\%$. There is also 1% chance that $P_N = 52 > 50\%$ and thus subjects will hold the N-lottery, whose expected payoff is $100 \times 52\%$, and so on. Therefore, the expected payoff for this baseline decision is:

$$\frac{1}{2}(100 \times \frac{1}{2}) + \frac{1}{100}(100 \times 51\%) + ... + \frac{1}{100}(100 \times 99\%)$$

After observing the first draw, subjects will report their belief of p_1, which I assume to be calculated by Bayes rule. Similar as the above case, there will be $p_1\%$ chance that $P_N < p_1$ and thus the subjects will hold their initial lottery, which as expected payoff of $100 \times p_1$. There is 1% chance that $P_N = p_1 + 0.01 > p_1$ and thus subjects will hold the N-lottery, whose expected payoff is $100 \times (p_1 + 0.01)$, and so on. Therefore, the expected payoff for the first draw decision is:

$$p_1(100 \times p_1) + \frac{1}{100}(100 \times (p_1 + 0.01)) + ... + \frac{1}{100}(100 \times (p_1 + 0.99 - p_1))$$

Again, after observing the second draw, if the subject is Bayesian, the expected payoff for making a Bayesian posterior(BU) is:

$$BU(100 \times BU) + \frac{1}{100}(100 \times (BU + 0.01)) + ... + \frac{1}{100}(100 \times (BU + 0.99 - BU))$$

If the subject is Average, the expected payoff for making an Average posterior(AU) is:

$$AU(100 \times AU) + \frac{1}{100}(100 \times (AU + 0.01)) + ... + \frac{1}{100}(100 \times (AU + 0.99 - AU))$$

For each round, each of the above decision may result as the payment for this round since it is determined by rolling a dice. In addition, the expected payoff for baseline decision and first draw decision are the same for both Bayesian updater and Average updater, therefore, the expected payoff for being Bayesian differs from that for being Average because they have different posterior belief after observing the second draw. So, if we want to find how Bayesian's expected payoff differs from Average's, we only need to compare the last decision's expected payoff.

$$\begin{aligned}
&\text{Expected Payoff for Bayesian} = \\
&= BU(100 \times BU) + \frac{1}{100}(100 \times (BU + 0.01)) + ... + \frac{1}{100}(100 \times (BU + 0.99 - BU)) \\
&= 100BU^2 + (BU + 0.01) + (BU + 0.02)... + (BU + 0.99 - BU) \\
&= 100BU^2 + (0.99 - BU) \times 100 \times BU + (0.01 + (0.99 - BU)) \times \frac{(0.99 - BU) \times 100}{2} \\
&= 50BU^2 - 0.5BU + 49.5
\end{aligned}$$

Expected Payoff for Average =

$$= BU(100 \times AU) + \frac{1}{100}(100 \times (AU + 0.01)) + ... + \frac{1}{100}(100 \times (AU + 0.99 - BU))$$

$$= 100AU^2 + (AU + 0.01) + (AU + 0.02)... + (AU + 0.99 - AU)$$

$$= 100AU^2 + (0.99 - AU) \times 100 \times AU + (0.01 + (0.99 - AU)) \times \frac{(0.99 - AU) \times 100}{2}$$

$$= 50AU^2 - 0.5AU + 49.5$$

If we set Expected payoff for Bayesian minus Expected payoff for Average > 0, after some derivation, we will get $BU+AU > 0.01$. This means as long as we have $BU+AU > 0.01$, being a Bayesian will have higher expected payoff in each round than being an Average. The only cases that we might have $BU + AU \leq 0.01$ are: (1) when both draws have 0 black balls or (2) there is one black ball in the first draw and 0 in the second draw.

The probability for (1) is:

$0.5 \times (\frac{6}{10})^{15}(\frac{6}{10})^{10} + 0.5 \times (\frac{4}{10})^{15}(\frac{4}{10})^{10} = 0.0001422\%$.

The probability for (2) is:

$0.5 \times (\binom{15}{1}(\frac{4}{10})(\frac{6}{10})^{14}) \times (\frac{6}{10})^{10} + 0.5 \times (\binom{15}{1}(\frac{6}{10})(\frac{4}{10})^{14}) \times (\frac{4}{10})^{10} = 0.001422\%$.

Therefore, the probability we have $BU + AU \leq 0.01$ is the probability for (1) + the probability for (2) = 0.0015642%, which means there is 99.998436% chance that being Bayesian will have a higher expected payoff than being Average. This could explain why at later stage of the game, subjects are less likely to behave like Average.

A.4 Experiment Instructions

A.4.1 Parm-A sessions

Experiment Instructions

Welcome! This is an experiment in the economics of decision-making. If you pay close attention to these instructions and make good decisions, you can earn a significant sum of money which will be paid in cash at the end of the session.

Please turn off your phone and other electronic devices. If you have any questions, or need assistance of any kind, please raise your hand and someone will come to you. We expect and appreciate your cooperation.

Basic Idea

There will be a lottery that will worth either \$100 or \$0 with equal probability.

You can either play the lottery or exchange it with another lottery that will be described below. Here is how it works: you decide the "chances out of 100" that the lottery you own worth \$100 and enter that probability into the computer. Then we will offer another lottery called N-lottery that pays the same amount \$100 with probability P_N and \$0 otherwise. The probability P_N is determined randomly between 0 to 99. If the probability you enter in the computer is higher than the probability P_N, then you will keep your initial lottery. If the probability you enter in the computer is smaller than the probability P_N, then you exchange your initial lottery with us and thus hold N-lottery.

For each lottery you own, you will enter your belief of the "chances out of 100" three times. For the first time, you don't know anything about this lottery other than it worth either \$100 or \$0 with equal probability, and you need to report your belief the "chances out of 100" that the initial lottery worth \$100.

For the second and third time, we will provide you some additional information about your initial lottery in order to help you guess the true value of your lottery so that you could adjust your belief the "chances out of 100" that the initial lottery worth \$100.

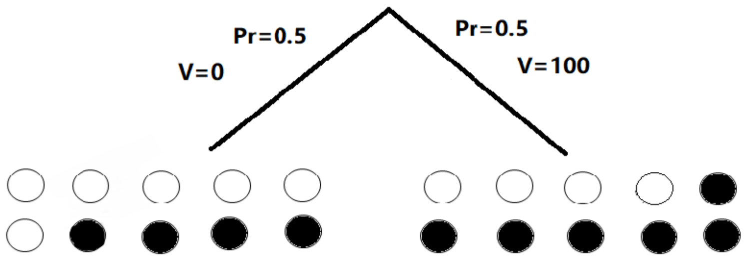


Figure A7: Prior and Urn Content (Parm-A)

The computer will help you guess the true value by drawing several balls from one of two bags. After the computer draws a ball, it replaces it in the same bag. As you can see in figure A7, if the true value is \$0, then the computer will draw balls only from the left bag, which contains 4 black balls and 6 white balls; if the true value is \$100, the computer will draw balls only from the right bag, which contains 6 black balls and 4 white balls.

In the first draw, the computer will draw 15 balls in the bag that corresponding to the true value of the lottery and it will tell you how many of the balls are black and how many are white.

Next, in the second draw, the computer will draw 10 balls in the same bag and will tell you how many of the balls are black and how many are white.

To summarize, you know at the beginning that the lottery will take value $100 and $0 with equal probability. But after seeing the first draw, you should have a better idea how likely this lottery worth $100 or $0 and thus you may adjust the probability that lottery worth $100. In particular, the more black balls and the fewer white balls you see, the more likely it is that the balls were drawn from the "good" $100 bag. Similarly, after seeing the second draw, you may want to further adjust the probability that the lottery worth $100.

Therefore, you will make 3 decisions in each round: 1. Decide the probability that the initial lottery worth $100 without seeing any draw. 2. Decide the probability that the initial lottery worth $100 after seeing the first draw. 3. Decide the probability that the initial lottery worth $100 after seeing the second draw.

What do you need to do?

What should you do in this experiment? You should enter the probability that is exactly equal to what you believe the initial lottery worth $100. Suppose you believe that this lottery takes value of $100 with probability 80/100, then the best thing you can do is to enter 80 into the computer! Why? Imagine you enter higher instead, say, 87. What could happen? For example, if the randomly drawn P_N for N-lottery is between 80 and 87, then you will NOT exchange your initial lottery and thus keep it. You are worse off in this case since the initial lottery worth $100 with probability 80/100 to you and you could have exchanged it with N-lottery, which worth $100 with a probability higher than 80/100.

If the random probability P_N is lower than 80 or above 87, you get the same result as if you'd set your belief at your true belief of 80.

Similarly, if you enter a probability that is lower than 80, say 75. What would happen? The randomly drawn probability P_N could lie between 75 and 80. In this case, you will exchange your initial lottery with us and thus hold N-lottery. Again, you are worse off, because your initial lottery worth \$100 with probability 80/100 and you just exchanged it to the N-lottery, which worth \$100 with probability that is lower than 80.

There is no circumstance in which offering a price not equal to your true value is to your advantage; it can only decrease your earnings or make no difference.

How do you get paid?

In addition to your \$7 U.S. base payment, each decision you make in the experiment will result in the possibility of earning money. You will play 50 rounds and make 3 decisions in each round. After the experiment, you will be paid in cash for one round that you played here today. The number of the round for which you are to be paid is determined randomly by the computer. The computer will draw a number from 1 to 50 with equal probability and the number been drawn means your payment is from that round.

Remember, you have made 3 decisions in each round.

Decision 1: you enter a probability to exchange your lottery without any draw.

Decision 2: you enter a probability to exchange your lottery after seeing first draw.

Decision 3: you enter a probability to exchange your lottery after seeing second draw.

To determine which decision matters, the computer will roll a single six-sided dice. If 1 dot comes, decision 1 will be the one that counts. If 2 dots or 3 dots come, then decision 2 will be the one that counts. If 4 or 5 or 6 dots come, then decision 3 will be the one that counts. That is to say, the probability that decision 1 will matter is 1/6. The probability that decision 2 will matter is 2/6. The probability that decision 3 will matter is 3/6.

Because you will not know in advance which round and decision will be selected for payment, you should treat each decision as though it will be the one for which you will be paid. Furthermore, because the decision for which you will be paid is randomly determined, nothing you do in the experiment will affect the decision number for which you are to be paid.

Your cash payment depends only on your pricing decisions and luck.

(Note: 1 dollar in the game worth 0.15 dollar in real world, all the dollar amount in the instruction are dollar in the game.)

A.4.2 Parm-B sessions

Instructions for Parm-B sessions are the same as Parm-A, the only difference is that the urn content and number of draws for each signal are different.

A.4.3 Quiz questions for the Presentation

Question 1: If you enter that the initial lottery worth 100 with probability of 85%. Then the computer randomly picked 60 for the N-lottery. What is your payment?

Answer: You keep your initial lottery and earn whatever that lottery turns out to worth.

Question 2: If you enter that the initial lottery worth 100 with probability of 60%. Then the computer randomly picked 77 for the N-lottery. What is your payment?

Answer: You will hold the N-lottery and earn whatever that N-lottery turns out to worth.

Question 3: The computer throw a dice and it's 3. Which one of the payment will be your payment for this round?

A. Baseline Payment.

B. Payment for first draw.

C. Payment for second draw.

Answer: B

A.4.4 Observations where subjects did not enter their beliefs into the computer

If a subject did not enter anything when timeout on each page, the computer will automatically enter a number for the subject. The duration for each page is long enough for the subjects to report their beliefs, if they still did not enter their beliefs into the computer, the computer would treat them as if the signal on the current page did not help them updating their beliefs and thus subjects are uncertain about their beliefs. Therefore, the computer would enter "50" for the subject in the baseline page if there is no answer when timeout and it would enter the baseline page's (resp. first draw page) answer to the first draw page (resp. second draw page) if there is no answer when timeout. Such "automatic answering" was well explained to the subjects and they understood that, if they did not report their beliefs when timeout, they would be treated as if they did not learn from the current signal.

A.4.5 Explanation for the Incentive Compatibility of Elicitation Tasks

In baseline cases where subjects don't observe any signal, they will be asked to enter their posterior belief of the chances out of 100 (denoted as R) that the lottery worth 100. The same question will be asked again when subjects receive signal 1 and signal 2. In this probability elicitation method, the computer will exchange the subjects' initial lottery with another lottery called N-lottery. The N-lottery also worth either 100 or 0, and the probability that it worth 100 is a random number that the computer draws from 0 to 99. Then the computer will compare the randomly drawn probability (denote as P_N) with the stated probability R. If $R > P_N$, the subjects will keep holding their initial lottery. If $P_N > R$, the subjects will exchange their initial lottery with the N-lottery.

	$R_{stated} > R_{true}$	$R_{stated} = R_{true}$
$P_N < R_{true}$	Initial lottery	Initial lottery
$P_N \geq R_{stated}$	N-lottery	N-lottery
$R_{true} \leq P_N < R_{stated}$	Initial lottery	N-lottery

Table A6: Incentive Compatible

Table A6 explains why such method is incentive compatible. To see this, suppose the subject truly believes that her initial lottery worth 100 with probability R_{true}, but she overstates her belief of $R_{stated} > R_{true}$. If the random probability P_N is smaller than R_{true}, or greater than R_{stated}, subjects' payoff would be the same as if she stated the true belief. However, if the random probability P_N lies between R_{stated} and R_{true}, the subject will not make the exchange so she will hold the initial lottery. However, the initial lottery worth 100 with probability only R_{true} to her, if she states the truth and exchange with the N-lottery, she could have been better off since the

N-lottery worth 100 with probability P_N and that is higher than R_{true}. Thus, subjects' best strategy is to truthfully report their posterior belief.

The logic for understate their belief is the same as above. To see this, in table A7, if $P_N > R_{true}$ or $P_N \leq R_{stated}$, they payoff would be the same as if the subject stated the true belief. However, if $R_{stated} \leq P_N < R_{true}$, the subject would exchange her initial lottery to the N-lottery. She is worse off in this case because the N-lottery actually is worth 100 with a lower probability than the initial lottery.

	$R_{stated} < R_{true}$	$R_{stated} = R_{true}$
$P_N > R_{true}$	N-lottery	N-lottery
$P_N \leq R_{stated}$	Initial lottery	Initial lottery
$R_{stated} \leq P_N < R_{true}$	N-lottery	Initial lottery

Table A7: Incentive Compatible

Appendix B

Chapter 2 Appendix

B.1 Supplementary Figures and Tables

B.1.1 Notes for Data Cleaning

- Some "miss-reported" data is corrected. One miss-reporting is no answer collected from some survey questions. Subjects might spent too much time thinking of their responses and didn't realize the time was out for the current page. Another miss-reporting is the answers are out of the reasonable range (e.g., the subject accidentally reported a probability of 120% for question 1). To deal with the "miss-reported" data, I used the midpoint of the ranges as their answers for pre-trading survey 1 because it is treated as if the subject "did not know what to enter for this question." I use the answers of the previous survey to replace the "miss-reported" answers in pre-trading survey 2 and the post-trading survey because it is treated as if the subject "did not know what to update for their beliefs."

- 1.88% (0.78%) of the observations from OBT(TBT) sessions are corrected in the above way.

- Simply omitting the above “miss-reported” observations is also performed, and both omitting and correcting the “miss-reported” data provide the same results.

- Two sessions’ data from OBT are not included in the analysis:

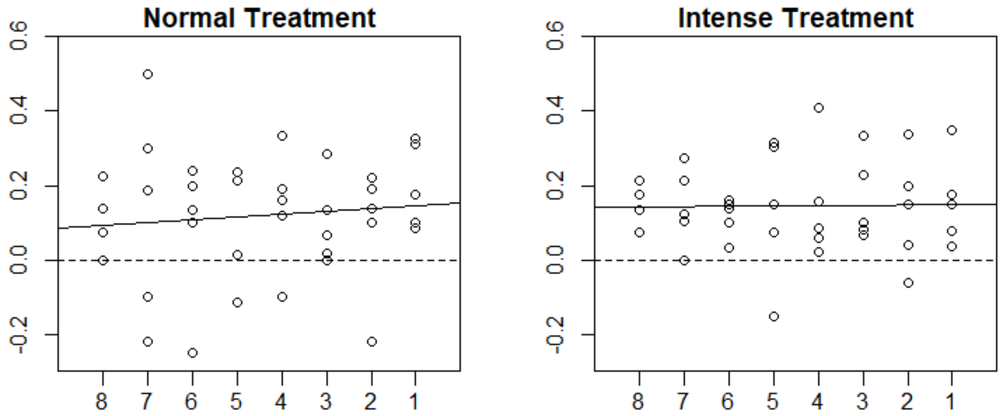
  - First, one of the sessions was canceled halfway due to an internet problem.
  - Second, session 11/03/2021’s (one intense treatment) data indicates two players did not understand the game well. One of them simply sold all his/her assets at low prices at the beginning of all the periods and did not do anything else. Another trader offers low asks regardless of his/her private signals. Each subject’s average earning from participating in the game is around \$21, but these two players only made \$12.
  - Qualitatively speaking (only look at the graphs), those graphs on belief updating and share holdings are similar with and without 11/03/2021’s data. But the graphs on market prices and trading behaviors are different if 11/03/2021’s data is included: most of the transaction prices in that session are lower than 200 while all other intense sessions’ prices are around 500, as illustrated in the paper. Bids and Asks’ deviation from RE-prices are significantly greater if that session’s data is included.
  - Therefore, we believe including these two sessions’ data does not help answer the research questions.

B.1.2 The Assessment of IQ Performance

Right after receiving the color signal, we expect most traders' elicited beliefs of the probability that the true state is G, which is denoted as p_1, to deviate from 0.5 in the treatments since traders are motivated. Let's denote $p_1 - 0.5$ as y. Therefore, green (red) traders are, on average, expected to have a positive (negative) y because green (red) traders are motivated to believe that state G (B) is more likely to be the true state. If they can accurately assess their IQ test performance, we would expect a positive (negative) correlation between y and their actual IQ ranks when traders are green (red). For example, if a trader's IQ rank is 1st (top), she will always be assigned to the H-group. Therefore, when she is green (red), she would immediately know that the true state is G (B) with high certainty, so her y would take a large positive (negative) value. Similarly, if a trader's IQ rank is 8th (bottom), she would know she is always in the L-group. So, if she is green (red), she would immediately know that the true state is B (G) with high certainty. However, the following correlations and graphs largely reject the above hypothesis.

Using data from OBT sessions, Figure B1 plots green traders' y (y-axis) and their actual IQ rankings (x-axis) in both normal and intense treatments. The first impression is that most traders' y is larger than zero regardless of their IQ ranks, indicating a positive bias when marked as green. Subjects with low IQ ranks tend to state similar positive y values as those with high IQ ranks. For example, on the left panel, one of the subjects got an IQ rank of 7th, but she is still confident about her IQ test performance and has a high y value that equals 0.5. Those who ranked top in the

IQ test in both panels do not necessarily have the highest y value. Graphically, we see slightly positive correlations. However, the Pearson correlation test suggests that the correlation is insignificant ($p = 0.5095$ for Normal and $p = 0.9193$ for Intense).



Note: Each observation is an average subject measure (e.g., given one trader's IQ rank, what is her average y when she was marked as green in a session). The following figures in this subsection are all subject average measures.

Figure B1: Green Traders' y and Actual IQ Ranks

Figure B2 plots red traders' y (y-axis) and their actual IQ rankings (x-axis). The patterns are reversed, but the suggested results are similar. For red traders, the correlation is expected to be negative. The negative correlations are not significant in both normal ($p = 0.2472$) and intense treatments ($p = 0.1571$). We can see from the graph that subjects who ranked top do not always have the lowest y. Traders with a low IQ rank are also confident that they are in the H-group and believe the state is more likely to be B when marked as red.

Instead of using their actual IQ ranks, the following two figures plot traders' y with their actual IQ scores, which take value in the range of [0, 32.5], for robustness

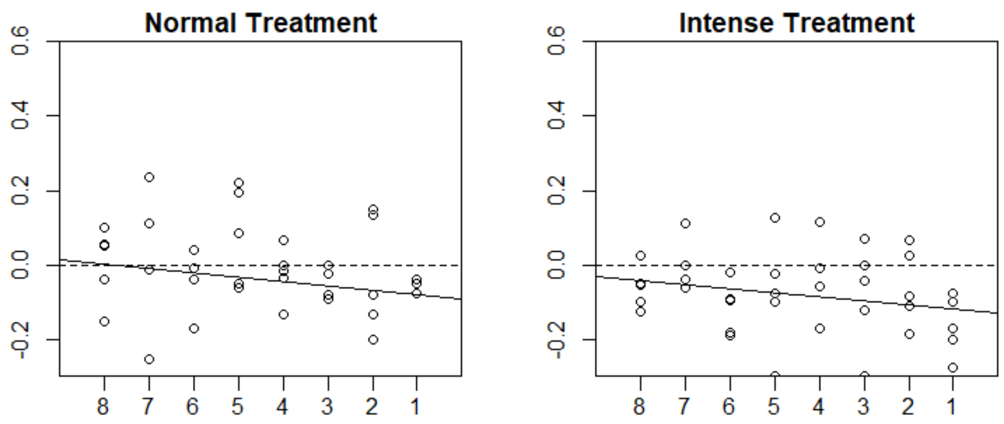


Figure B2: Red Traders' y and Actual IQ Ranks

check. The results are similar. None of the correlations are significant.[1] In addition, we continue to reject significant correlations using data from TBT sessions (See P-values of Pearson Correlation Test in Table B1).

	IQ Rank		IQ Score	
	Normal	Intense	Normal	Intense
Green	0.3532	0.901	0.9041	0.931
Red	0.0231	0.8495	0.2418	0.7674

Table B1: P-values for Pearson Correlation Test (TBT Sessions)

The results in this analysis suggest that traders' elicited beliefs of p_1 are not significantly correlated with their actual IQ performance. Traders do not exhibit accurate assessments of how well they did on the IQ test. Therefore, the deviation of elicited p_1 from 0.5 is thus a bias due to IQ and payoff motivation.

[1]The p-values are 0.3275 (0.8130) for Normal and 0.7456 (0.11) for Intense when traders are green (red).

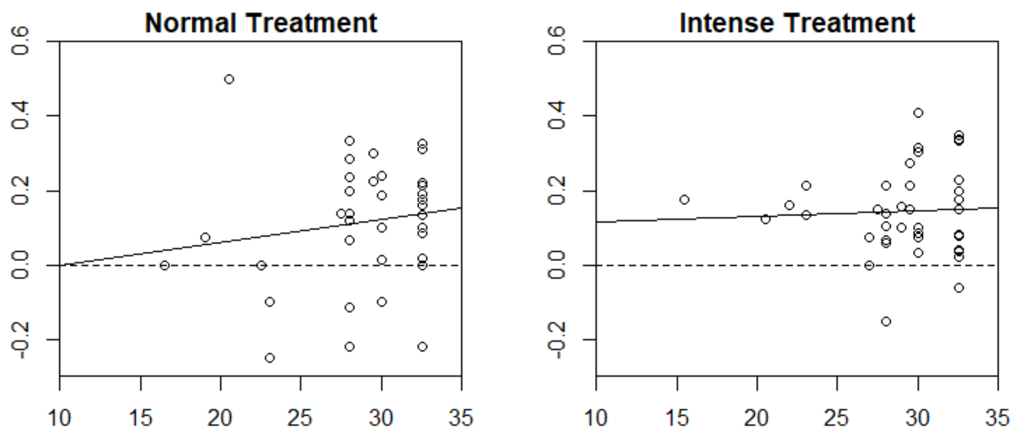


Figure B3: Green Traders' y and Actual IQ Scores

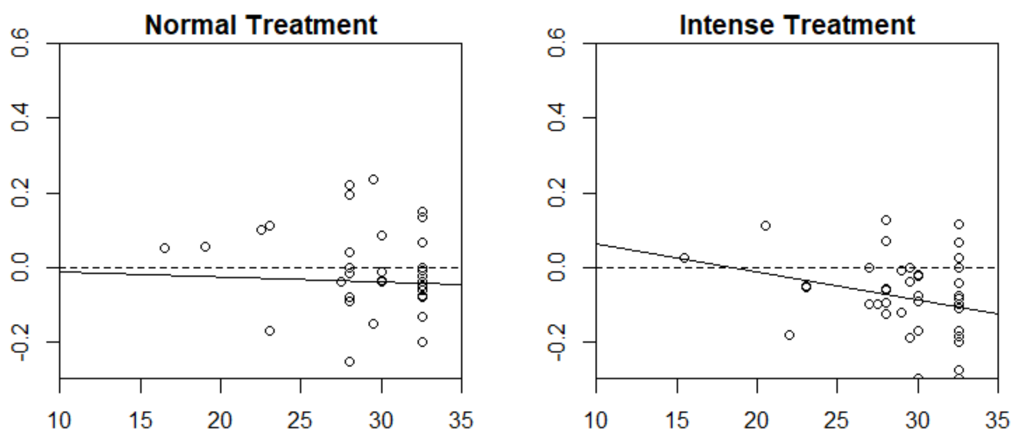
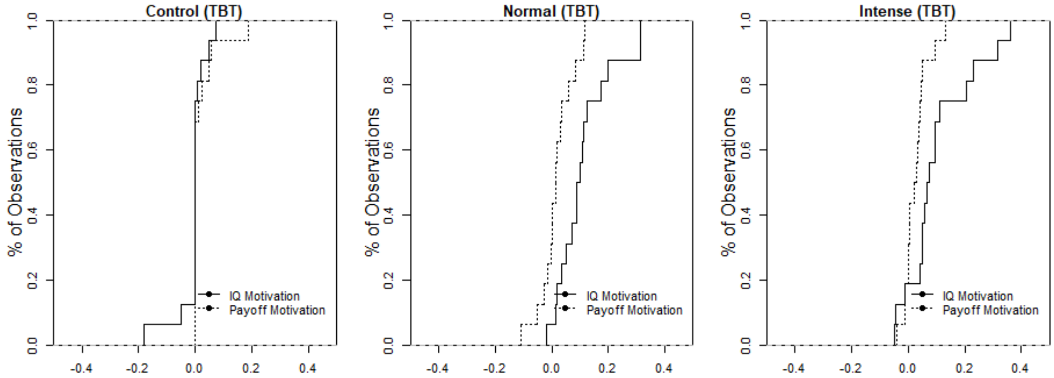


Figure B4: Red Traders' y and Actual IQ Scores

B.1.3 IQ Motivation and Payoff Motivation (TBT)



Note: This Figure replicates Figure 2.4 but using TBT sessions data. Each obervation is a subject average measure. There are 16 traders in each treatment.

Figure B5: IQ Motivation vs. Payoff Motivation

	TBT		
	Control	Normal	Intense
Average IQ Motivation	-0.005 (0.05)	0.11 (0.10)	0.10 (0.12)
Average Payoff Motivation	0.02 (0.05)	0.02 (0.06)	0.03 (0.04)

Note: Standard deviations are in the parentheses.

Table B2: Average IQ Motivation and Payoff Motivation

B.1.4 Belief Updating for Question 2 in OBT

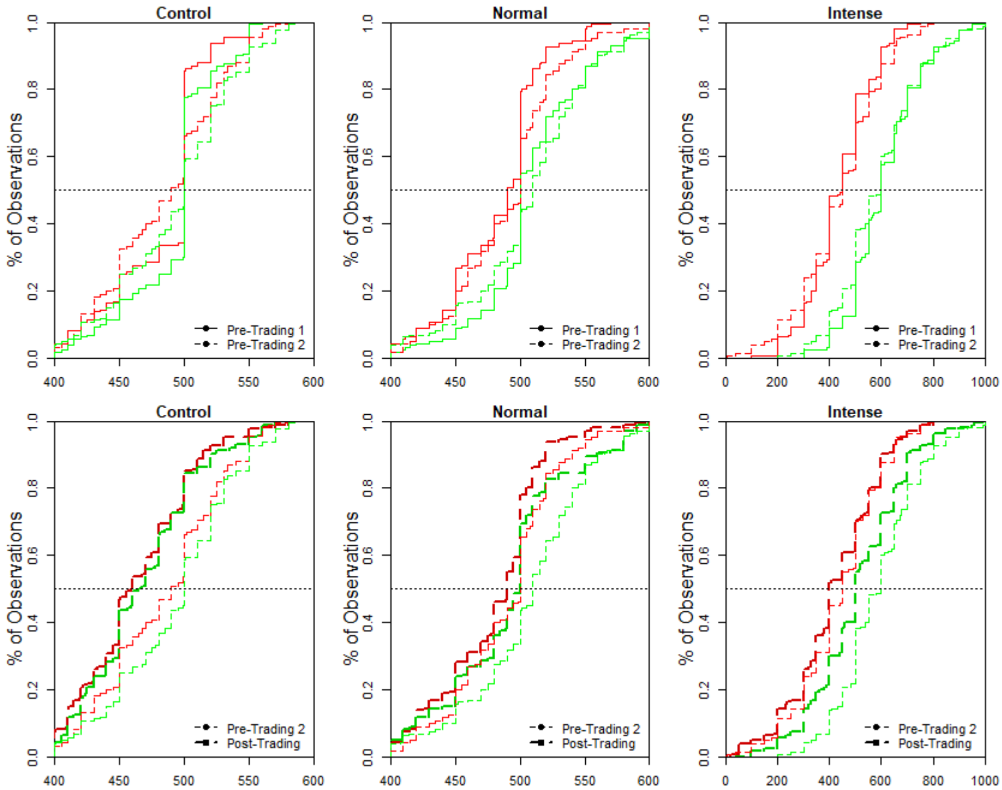
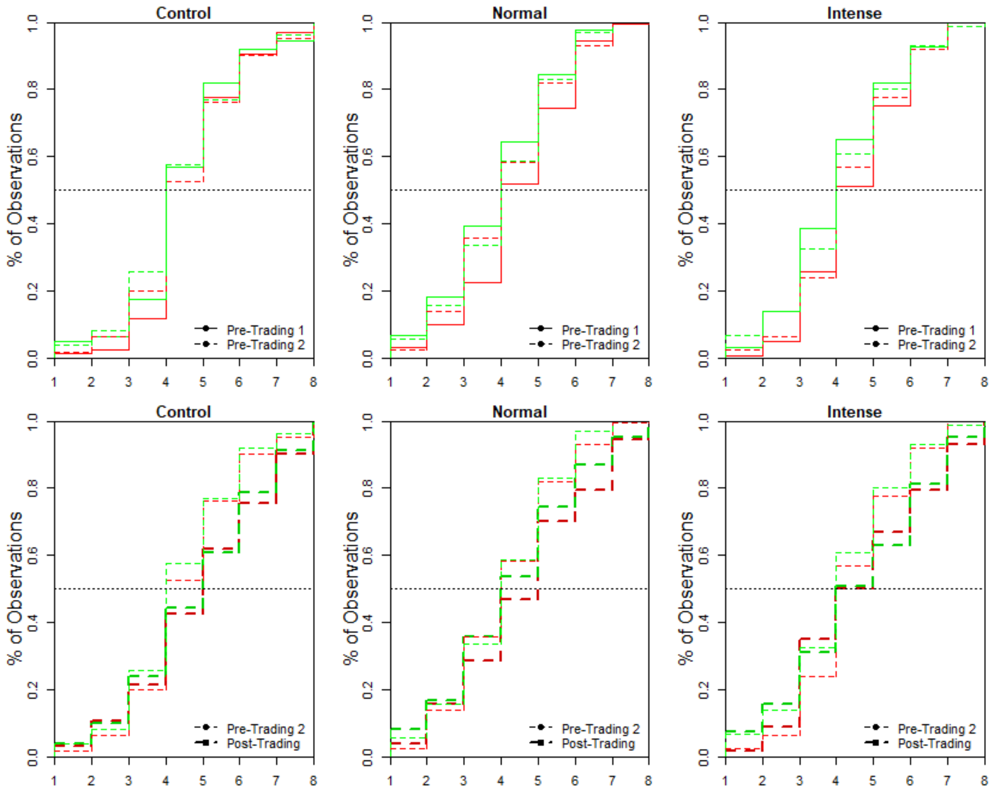


Figure B6: Belief Updating (Question 2, OBT)

B.1.5 Belief Updating for Question 3 in OBT



Note: This figure replicates figure 2.5 with elicited beliefs of question 3 to compare traders' beliefs on profit ranking among all 8 traders at different survey pages across treatments. The x-axis takes a value from 1 to 8 since there are 8 traders in each session.

Figure B7: Belief Updating (Question 3, OBT)

B.1.6 Belief Updating for Question 1 in TBT

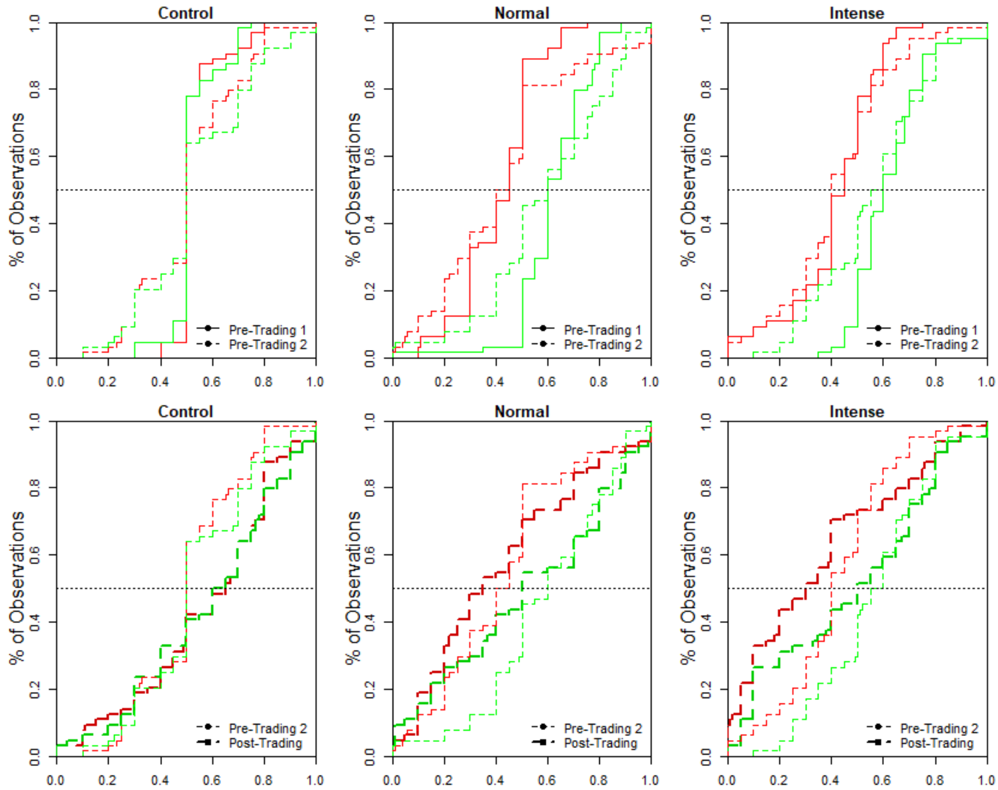


Figure B8: Belief Updating (Question 1, TBT)

B.1.7 Belief Updating for Question 2 in TBT

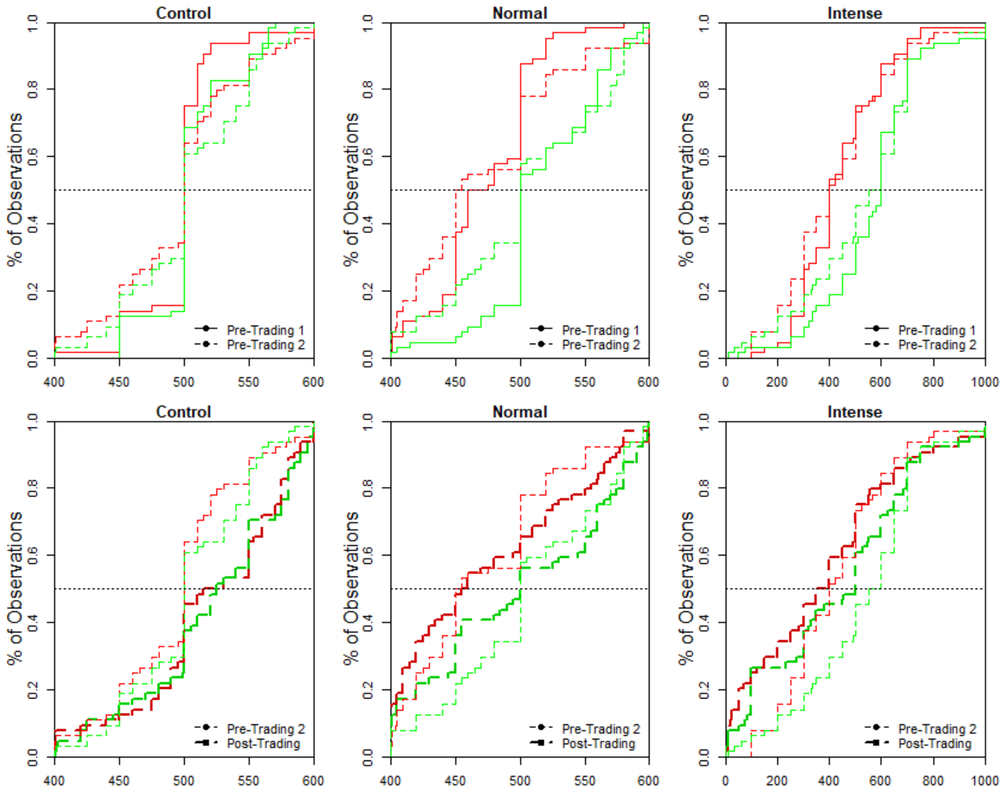


Figure B9: Belief Updating (Question 2, TBT)

B.1.8 Belief Updating for Question 3 in TBT

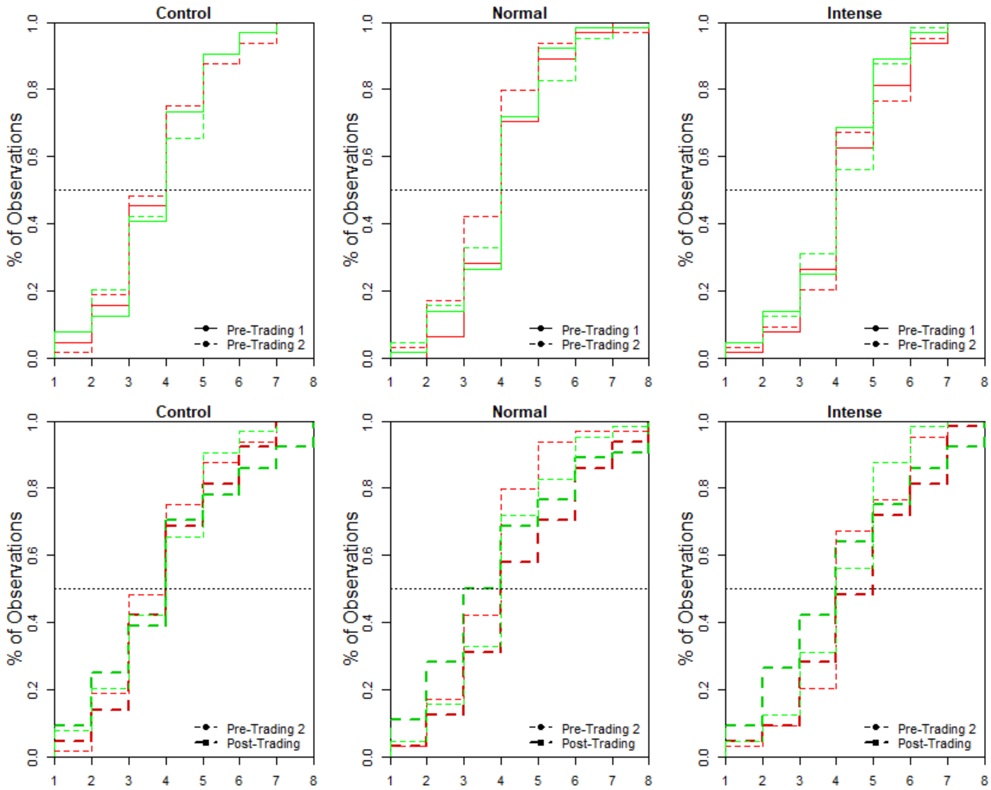
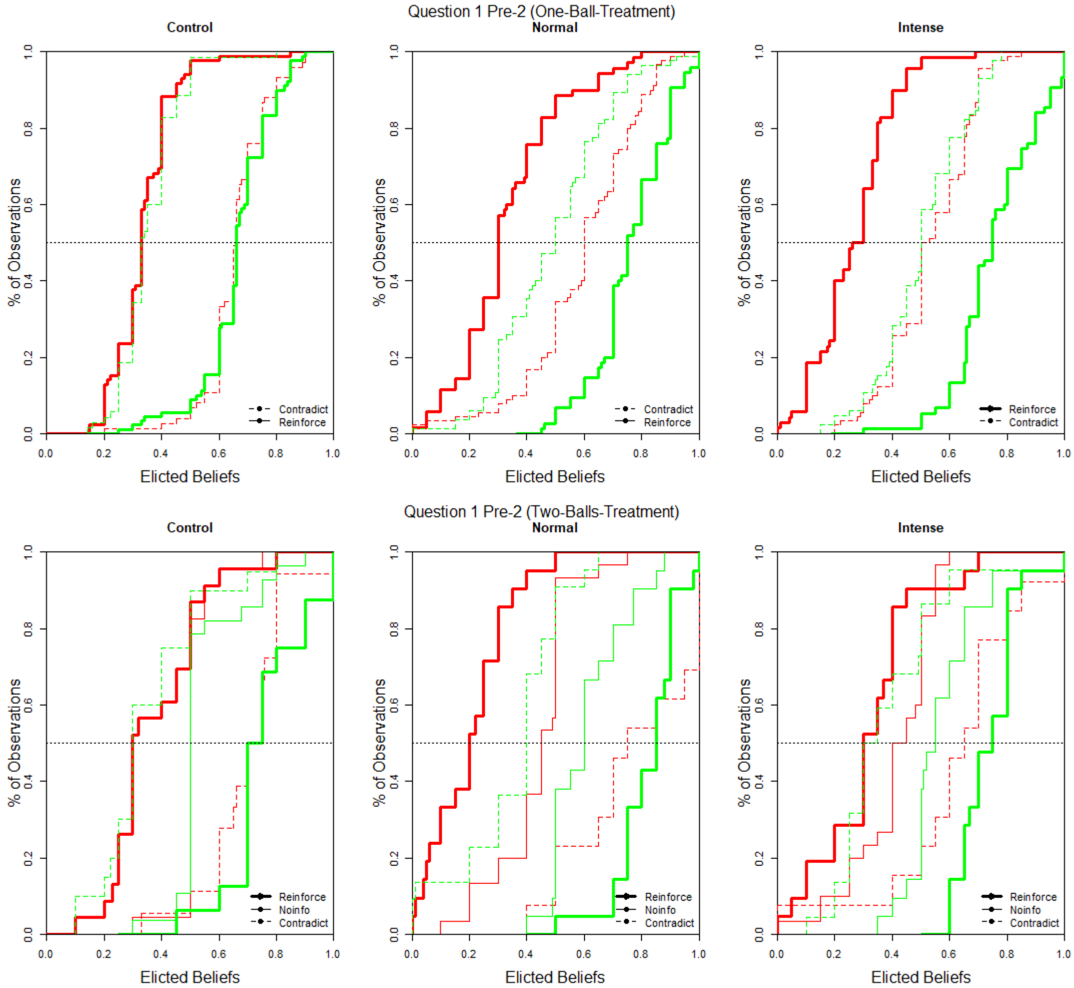


Figure B10: Belief Updating (Question 3, TBT)

B.1.9 Traders' Beliefs Given Contradicting Ball Signal



Note: In the control of TBT sessions, the solid red line and solid green line largely overlap each other; this is why it is hard to observe the solid red line. When traders are in this control treatment, receiving one black ball and one white ball does not provide much information to the traders, and thus the elicited beliefs from green and red traders are largely at 50%.

Figure B11: Beliefs of Question 1 Given Contradicting Ball Signal

The above graphs show elicited point probability beliefs of question 1 as CDFs when traders face a ball signal that contradicts, reinforces, or provides no information (no info) to their IQ motivation. A ball signal is called "contradict" when green (red) traders receive the white (black) ball(s) since the color of the ball indicates the true state is more likely to be the opposite of their preferred state. "Reinforce" means the the ball's color indicates the same state as traders' IQ motivations. No information (No info) means the realized drawn balls contain one black ball and one white ball, and this case only applies to TBT sessions. The upper (bottom) three graphs show us the CDFs in OBT (TBT) sessions. Green (red) color refers to elicited beliefs from green (red) traders. The thick solid line refers to traders who received the "reinforce" ball signal, the dashed line refers to traders who received the "contradict" signal, and the solid line for TBT sessions refers to traders who received the "noinfo" ball signal.

In control with no polarized motivated beliefs, ball signals do not contradict or reinforce traders' beliefs; therefore, we could see that traders who receive the same realized ball signal report similar point probabilities. In the normal and intense treatment, both OBT and TBT sessions show that traders who got contradicting signal adjust their beliefs towards the state they do not prefer, suggesting that a contradicting private signal has a significant impact on traders' belief adjustment. In addition, when motivated traders receive a private signal that indicates the true state is more likely to be Bad (i.e. two white balls), green traders' elicited beliefs, plotted as the green dashed line, first-order-stochastically dominate red traders' elicited beliefs, plotted as the thick red solid line. This pattern suggests that the polarized motivated beliefs still significantly impact traders' beliefs. The same pattern can also be found when motivated traders receive a private signal that indicates the true state is more likely to be Good.

In normal and intense treatment under TBT sessions, the "noinfo" signal does not provide any information for the traders to guess which state is more likely to be true. This is why the graph shows green traders' elicited beliefs first-order-stochastically dominate that of the red traders since the patterns of the CDFs are shaped by traders' polarized motivated beliefs.

Figure B12 replicates figure B11 using elicited beliefs of question 2. The patterns and the results are similar.

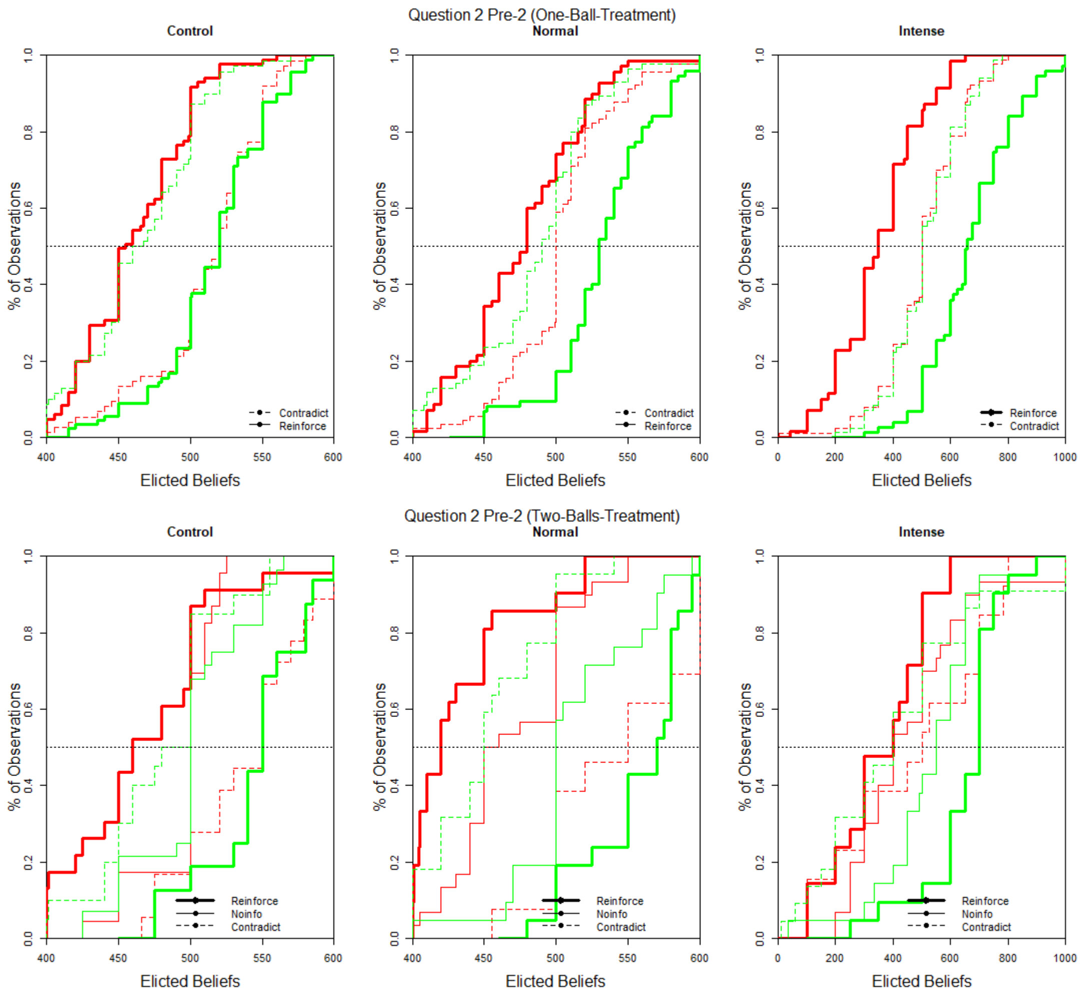
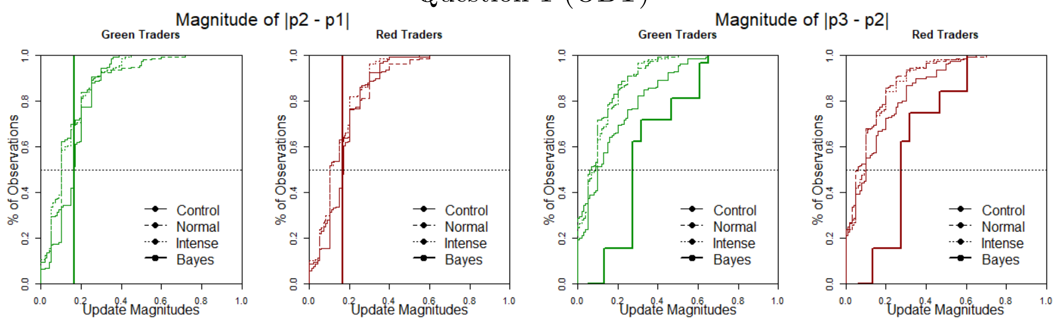


Figure B12: Beliefs of Question 2 Given Contradicting Ball Signal

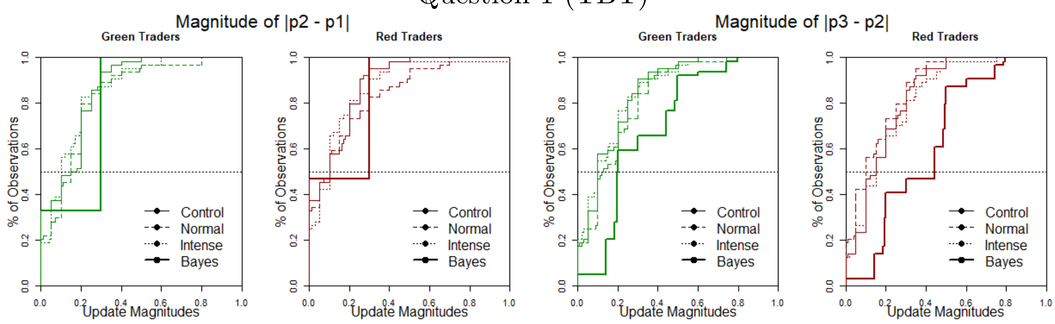
B.1.10 Belief Updating Magnitudes

This subsection tries to answer the following question: are traders' updating magnitudes (after receiving the ball signal and the trading) smaller when they hold polarized motivated beliefs because the beliefs are more persistent? Here, two update magnitudes are defined and measured: (1) How much do traders update their beliefs after receiving the ball signals (p_2) compared with their beliefs after the color signal (p_1), i.e. $|p_2 - p_1|$; and (2) how much do they update beliefs after trading (p_3) compared with p_2, i.e. $|p_3 - p_2|$.

Question 1 (OBT)



Question 1 (TBT)



Note: The upper (bottom) graphs show the two update magnitudes for question 1 in OBT (TBT) sessions.

Figure B13: Belief Updating Magnitudes (Question 1)

Figure B13 provides graphical evidence for the above questions (magnitudes for Question 1 only). The upper graphs in figure B13 plot $|p_2 - p_1|$ and $|p_3 - p_2|$ in OBT sessions as cumulative distribution function (CDF), where x-axis is the update magnitudes and the y-axis is the percentage of observations. The solid line, dashed line, and dotted line represents control, normal, and intense treatments, respectively. The update magnitude from perfect Bayesian is also presented in the graph as the thicker solid line. The CDFs of normal and intense treatments are close to each other, suggesting that traders with polarized motivated beliefs update their point probability beliefs after the ball signal and trading in a similar magnitude regardless of the financial stakes. For the control sessions, however, the CDF first-order-stochastically dominates that of the normal and intense for $|p_3 - p_2|$, indicating traders adjust their beliefs more after the trading when they do not hold polarized motivations. For $|p_2 - p_1|$, although the CDFs appear close to each other across three treatments, we can still find the control sessions first-order-stochastically dominate that of the normal and intense for the most proportion of the CDF, suggesting that traders in control adjust their beliefs more when they receive ball signals. The WSR tests support the graphical evidence: when compare the CDF of control with the CDFs of the other two treatments, the difference is statistically significant; while the difference is not significantly if we compare the two CDFs only between normal and the intense treatments.

The bottom graphs of figure B13 replicate the upper graphs with elicited beliefs from TBT sessions. The pattern states that the difference between all three CDFs is not substantial. The WSR tests indicate that the difference between any two CDFs, regardless of the treatments, is not statistically significant. This suggests a more precise ball signal improves motivated traders' overall update magnitudes to the signal and transaction prices so that their beliefs are less persistent.

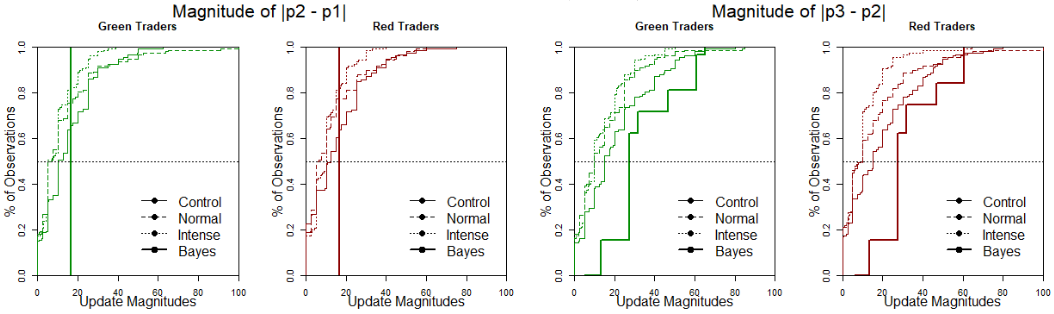
Since traders' willingness to accept is positively correlated with their point beliefs from question 1. Thus, the patterns of question 2 are similar and are provided in figure B14.

In addition, regardless of which question and magnitude it was measured, traders' updating magnitudes are, for the most part, lower than that of the perfect Bayesian, especially the magnitude of $|p_3 - p_2|$. This suggests traders exhibit non-Bayesian belief updating in terms of magnitudes and their magnitudes are smaller than Bayesian. This finding is consistent with many previous non-Bayesian belief updating literature:[2] The inverse-S-shaped belief updating pattern suggests that when Bayesian predicts a high(low) probability, subjects tend to report a relatively low(high) subjective probability that is closer to 0.5.

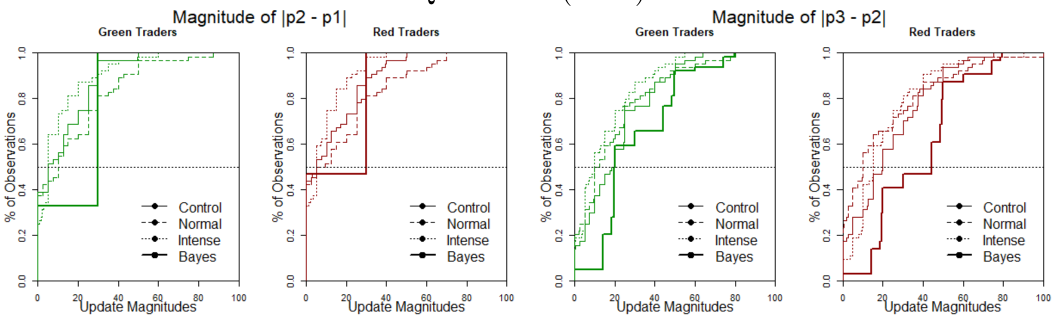
Given the above analysis, one additional conclusion can be made: An increased financial stake does not impact traders' belief updating magnitude. When only one ball is drawn from the bag as the private signal, traders hold no polarized motivated beliefs exhibit significantly higher update magnitudes, suggesting the motivated traders' beliefs are more persistent and do not adjust much when receiving new information from the market. However, increased precision of the private signal largely eliminates the differences, making traders across all treatments exhibit similar update magnitudes.

[2]See Holt and Smith (2009), Tversky and Kahneman (1992), Kahneman and Tversky (2013), for examples.

Question 2 (OBT)



Question 2 (TBT)



Note: Two updating magnitudes are plotted as CDF: (1) $|p_2 - p_1|$; and (2) $|p_3 - p_2|$. Solid CDF, dashed CDF, and dotted CDF means control, normal, and intense treatments, respectively. As a referene, the thick lines refers to the upate magnitudes from a perfect Bayesian's perspective. The upper (bottom) graphs show the two update magnitudes for question 2 in OBT (TBT) sessions.

Figure B14: Beliefs Updating Magnitudes (Question 1)

B.1.11 Shareholdings and the Price Spread

From section 2.3.3, we learned that traders who are motivated to prefer G(B) state hold significantly more(less) assets when the market closes in OBT. The graphs of final asset allocations under TBT are presented in figure B15, and the patterns are similar.

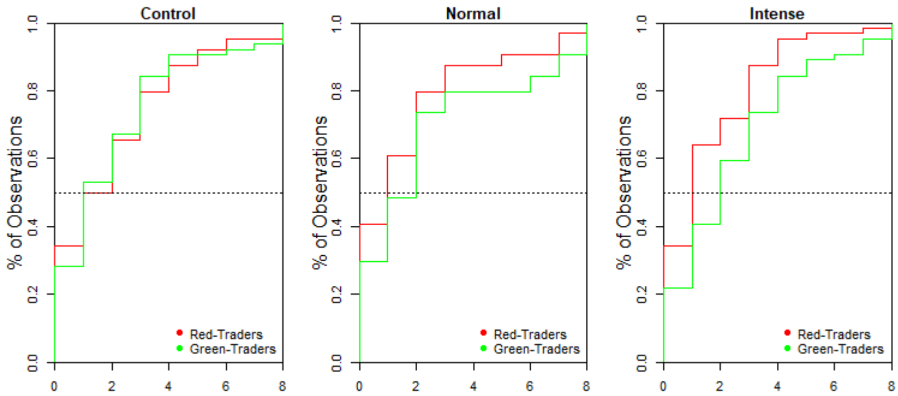
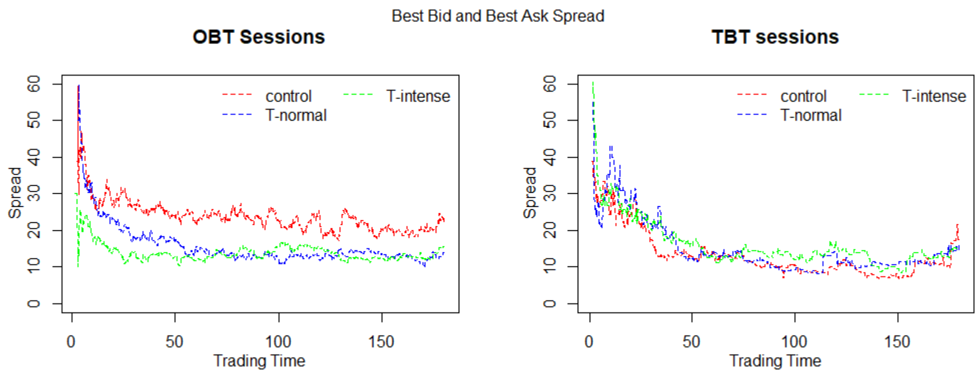


Figure B15: Share Holdings (TBT)



Note: The spread is calculated as the best ask minus the best bid in the market. The control and normal treatments have different dividends than the intense treatment, so the spread is rescaled to 100 for all treatments. Each trading period lasts 180 seconds, so the lines on the graph are the average of all the spread at each 0.1 seconds across different sessions within each treatment. The red, blue, and green dashed line refers to the control, normal and intense sessions. The left(right) figure is from OBT(TBT) sessions.

Figure B16: Spread

B.1.12 Market Prices Control-Treatment (OBT)

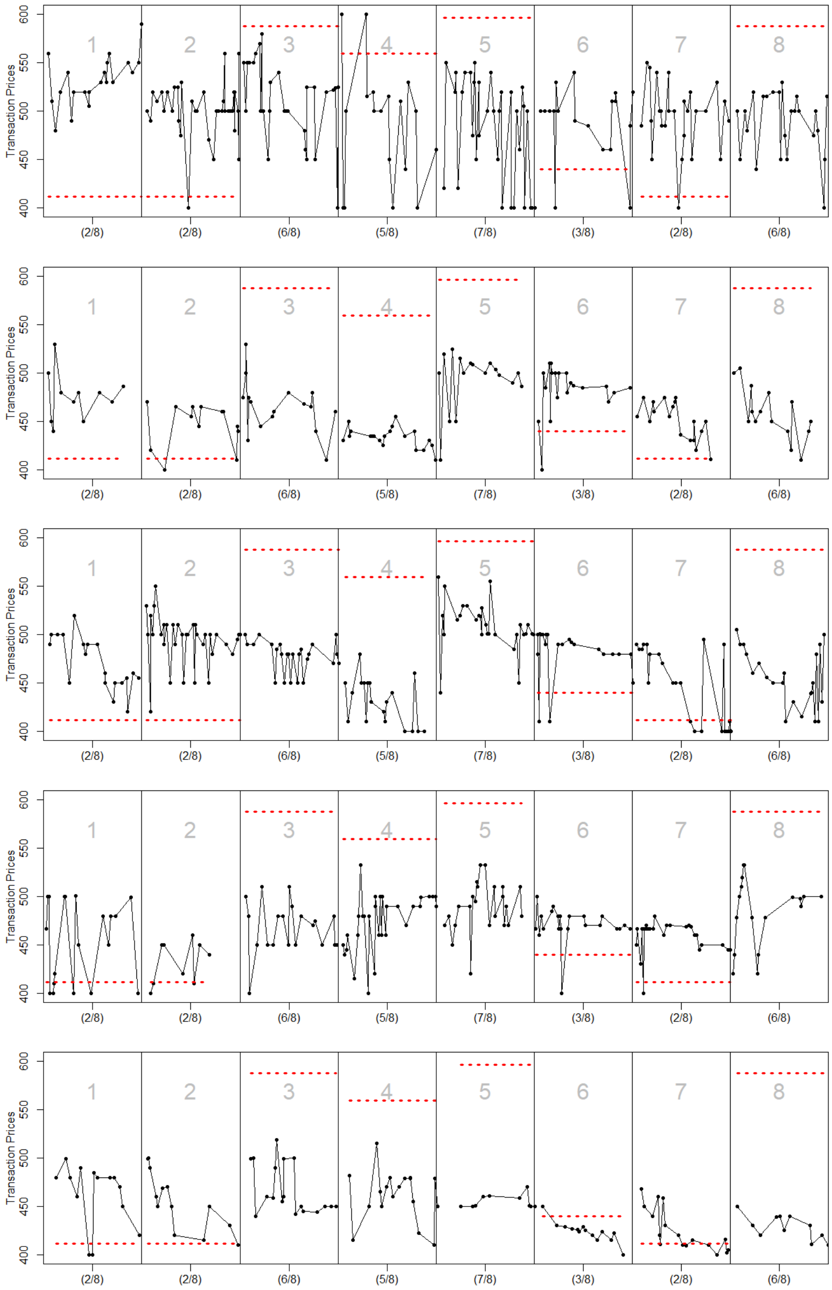


Figure B17: Transaction Prices in Control (OBT)

B.1.13 Market Prices Normal-Treatment (OBT)

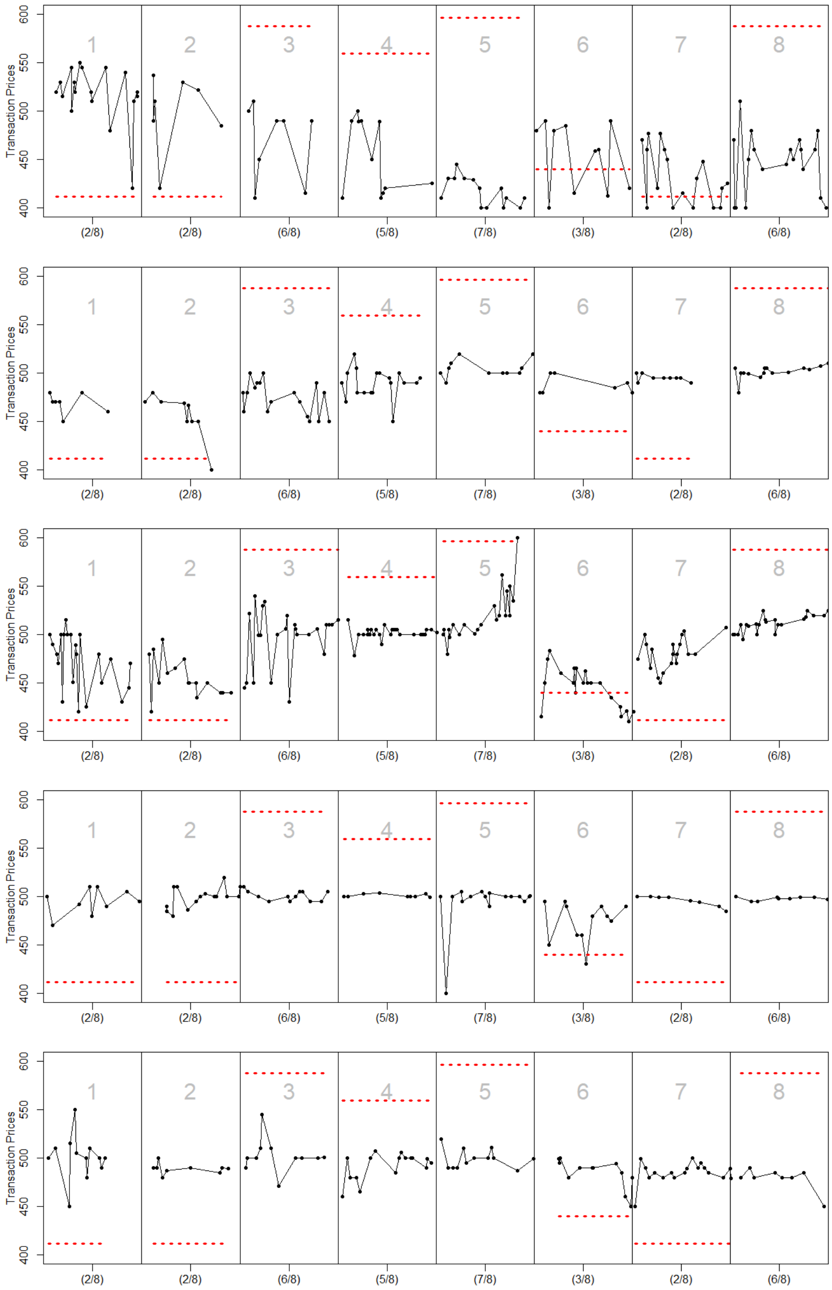


Figure B18: Transaction Prices in Normal (OBT)

B.1.14 Market Prices Intense-Treatment (OBT)

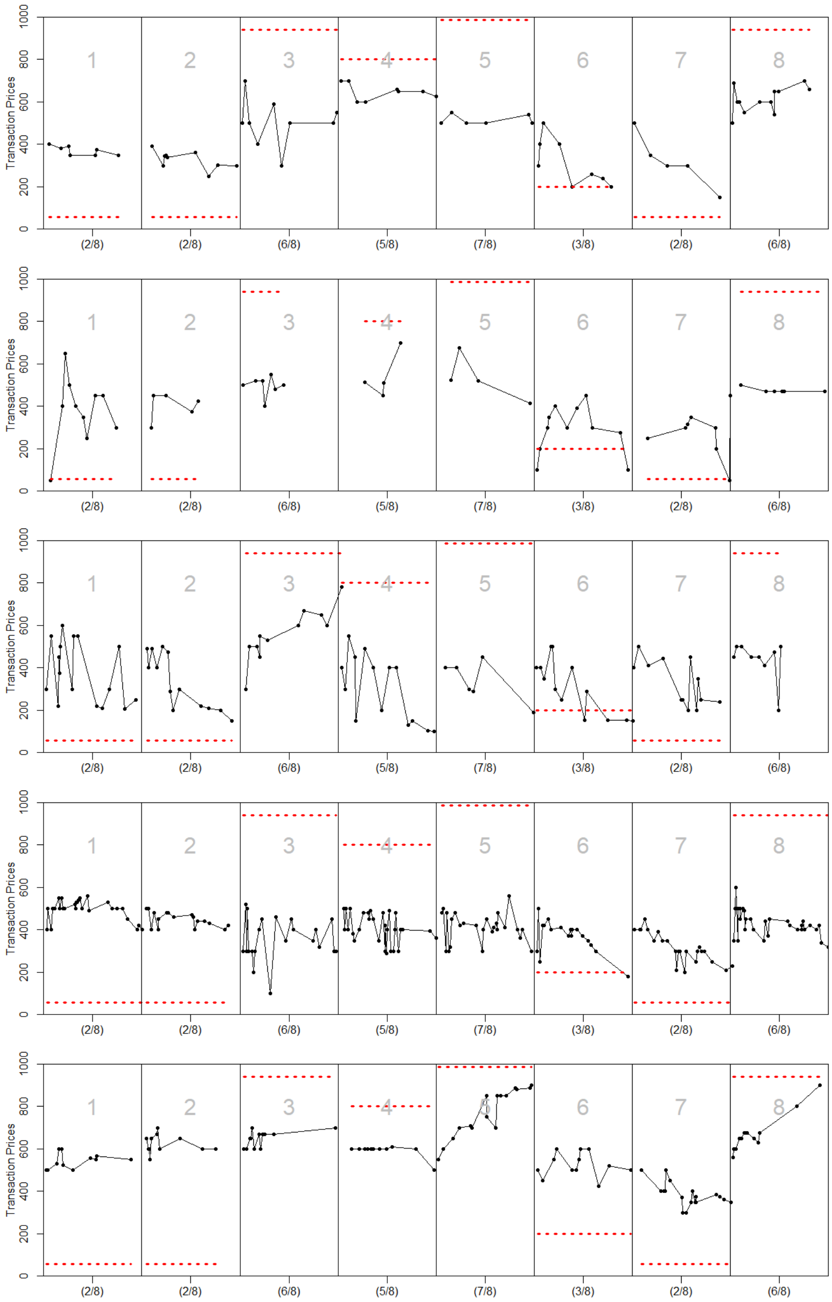
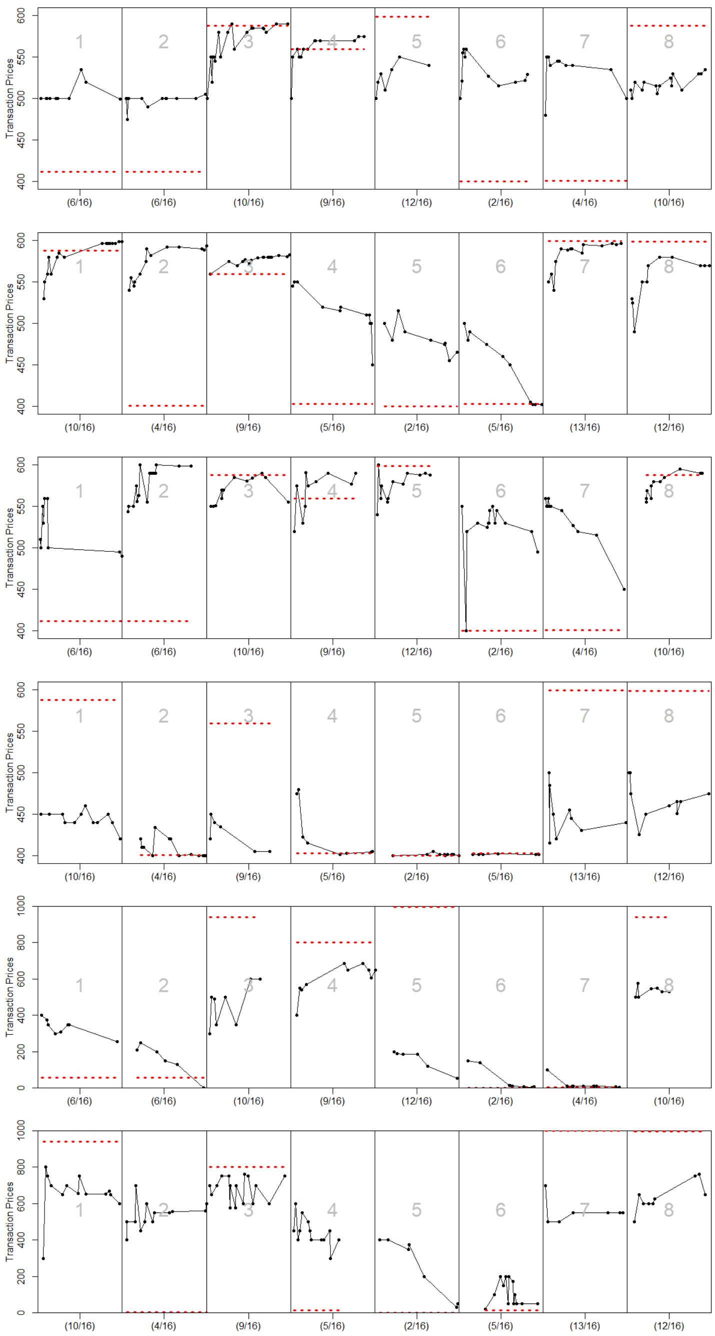


Figure B19: Transaction Prices in Intense (OBT)

B.1.15 Market Prices (TBT)



Note: The first two graphs are the control-treatment. The third and fourth graphs are normal-treatment. The last two graphs are intense-treament.

Figure B20: Transaction Prices in TBT

B.1.16 Regression (TBT)

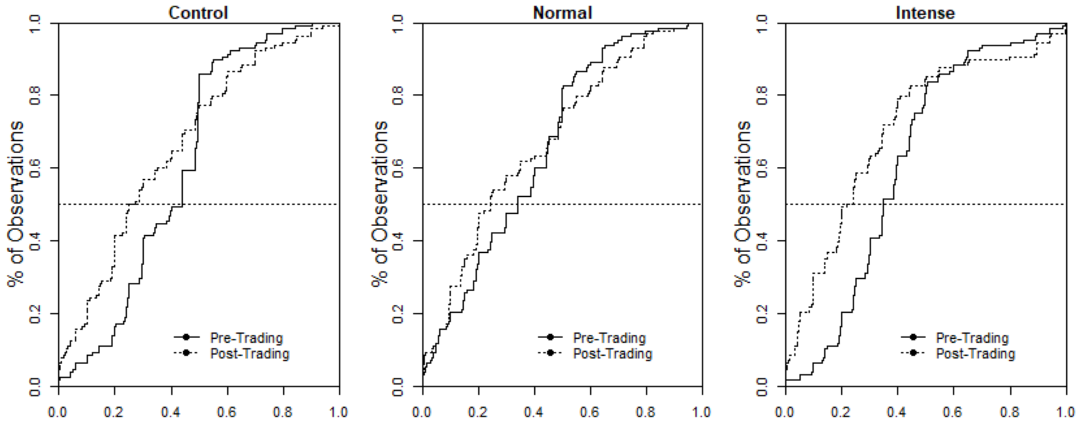
VARIABLES	(1) P-Error (Last Price)	(2) P-Error (Last Two Prices)	(3) Spread
β_C	24.59***	25.51***	9.982
	(5.358)	(4.695)	(13.76)
β_N	-4.658	-2.720	5.339
	(6.272)	(5.520)	(14.09)
β_I	2.303	0.112	-2.410
	(6.820)	(6.120)	(13.92)
Observations	48	48	48
R-squared	0.564	0.563	0.575

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: This Table replicates the regression results in Table 2.3 using data from TBT sessions. None of the estimated coefficients on column three is significant. This might be because the more precise ball signal (perhaps also traders' experiences) improves traders' valuations of the asset to reach a consensus over time, regardless of payoff motivation or IQ motivation.

Table B3: Regressions of Convergence

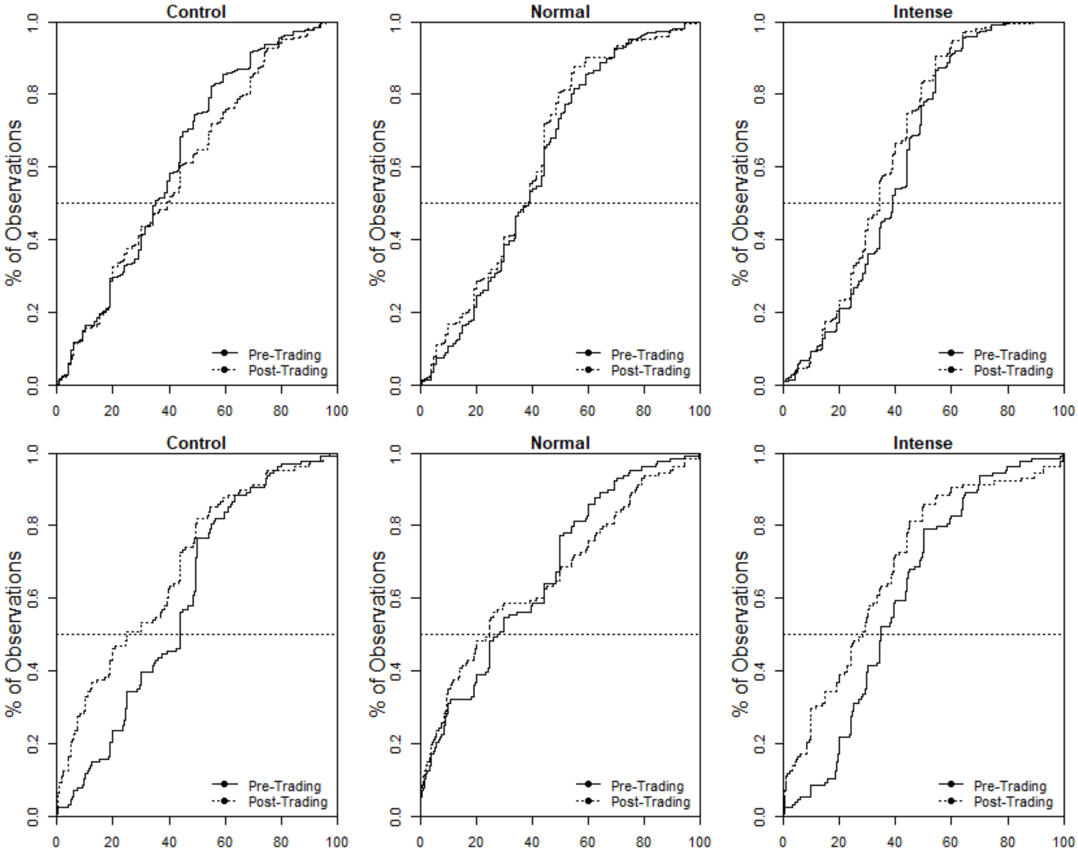
B.1.17 Belief Errors TBT



Note: This Figure replicates the analysis of section 2.3.4 using data from TBT sessions.

Figure B21: Belief Errors (Question 1, TBT sessions)

B.1.18 Belief Errors Using Question 2 and Question 3



Note: This figure replicates figure 2.8 using elicited beliefs of question 2. All the belief errors are rescaled to 100 for all treatments.

Figure B22: Belief Errors (Question 2)

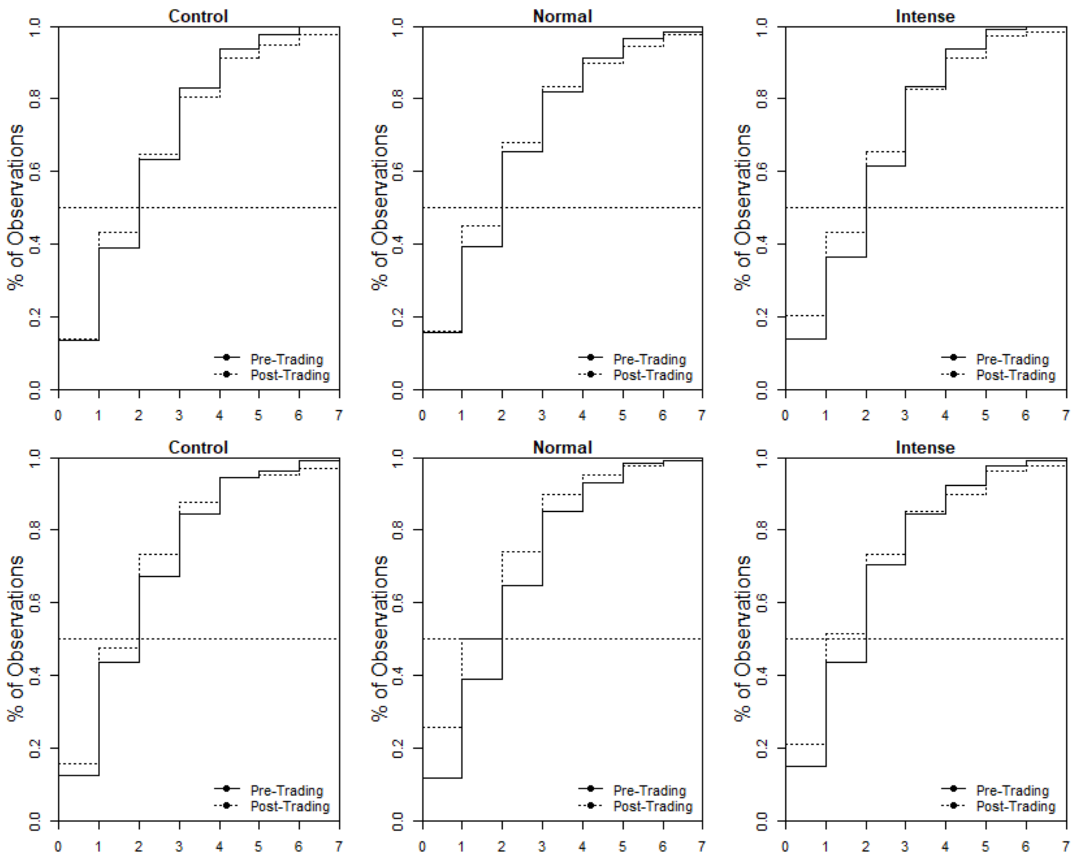


Figure B23: Belief Errors (Question 3)

This figure replicates figure 2.8 using elicited beliefs of question 3. The errors are calculated by taking the absolute difference between traders' reported beliefs and their actual profit ranking. From the above graphs and the WRS test, the results show that trading does not significantly impact traders' belief errors regarding their profit ranking.

B.1.19 Belief Errors with Directions

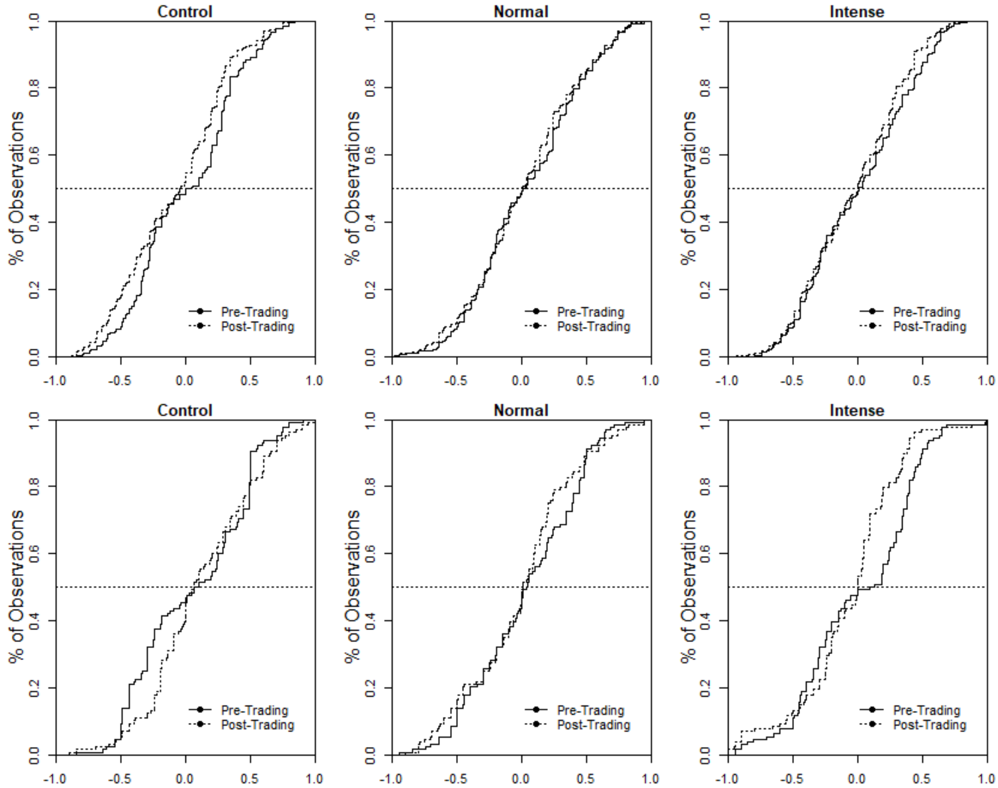


Figure B24: Belief Errors with Directions (Question 1)

This figure replicates figure 2.8 without taking the absolute value of the belief errors. The upper graph uses data from OBT sessions, and the p-values of the WRS test are 0.002, 0.554, and 0.2396 for control, normal, and intense, respectively. The bottom graph uses data from TBT sessions; the p-values for control, normal, and intense are 0.208, 0.368, and 0.112. The hypothesis test results suggest that when traders hold more precise signals under TBT, they do not significantly reduce their belief errors after trading, regardless of the treatment. Polarized motivations and intense financial stakes also do not significantly impact belief errors under both OBT and TBT sessions. In control under OBT, however, trading seems to shift belief errors towards negative values. When belief errors before trading take positive (negative) value, observations are predominately from the Bad (Good) state; this is because perfect Bayesian is low (high) when the state is B (G) and elicited beliefs are closer to 0.5 relatively. From transaction price dynamics (see section B.1.12) in control under OBT, we can see that prices have a relatively better convergence to RE-price when the state is B. But prices are, in general, lower than Bayesian prediction when the state is G. This explains why belief errors are closer to zero after trading when the state is B; it is because the price dynamics are more in line with traders' private signals. While in the good state, price dynamics are less consistent with traders' private signals, and thus they are less certain about the true state and report their beliefs closer to 0.5, resulting in a left shift of belief errors.

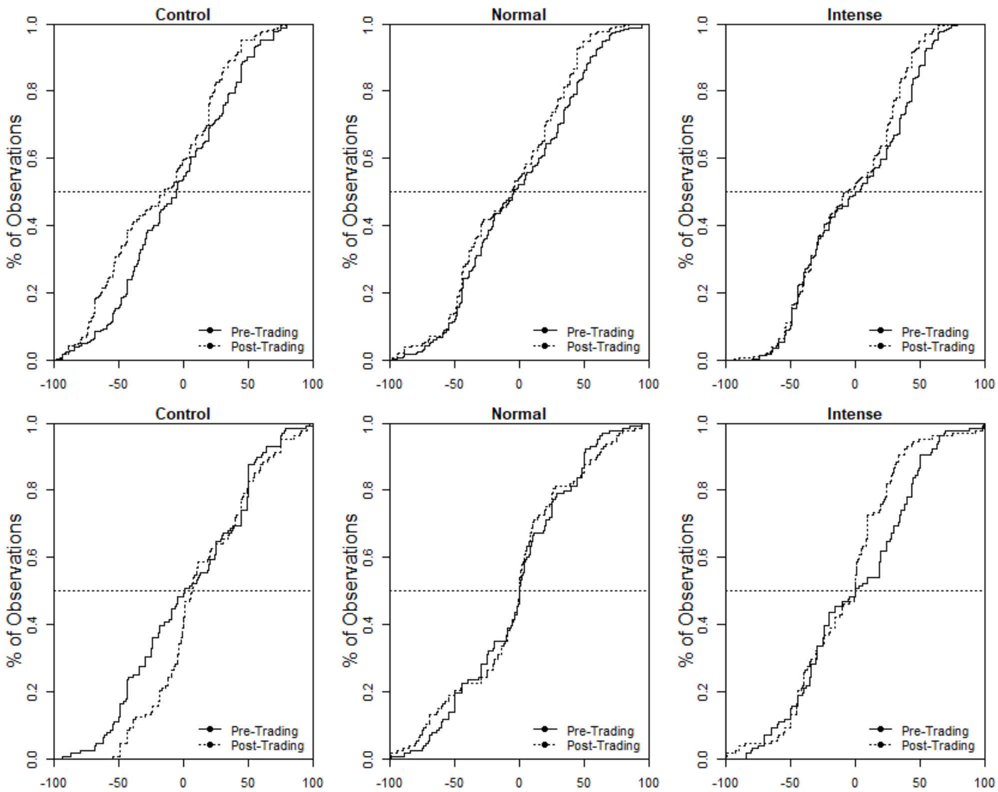


Figure B25: Belief Errors with Directions (Question 2)

Figure B25 uses data from question 2 and the patterns are similar. Figure B26 uses data from question 3 and we do not observe significant change in belief errors regardless of the treatment.

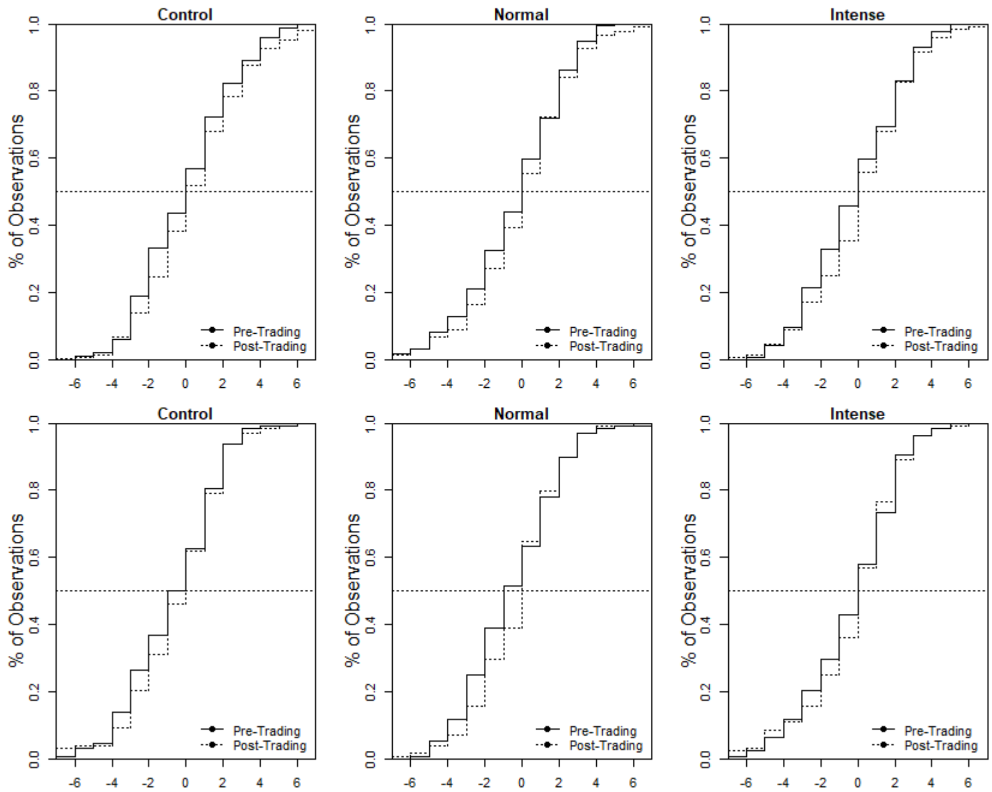


Figure B26: Belief Errors with Directions (Question 3)

B.1.20 Updating In Response to the Ball Signals

This subsection aims to find out how individual traders' responsiveness to the ball signals differs with and without polarized motivated beliefs and how drawing two balls from the bag affect responsiveness. The Bayes' rule in the odds ratio form is adapted from previous literature[3] to analyze traders' responses:

$$log(\frac{p}{1-p}) = log(\frac{p_0}{1-p_0}) + log(\eta_s),$$

where p_0 is the prior, p is the posterior, and η_s is the likelihood ratio of observing a realized signal, for example, in the OBT sessions, η_s of observing a black ball equals to $\frac{q}{1-q} = \frac{2/3}{1-2/3} = 2$. To estimate traders' responsiveness to different realizations of drawn balls in the OBT sessions, the following regression is applied:

$$log(\frac{p_2}{1-p_2}) = \alpha log(\frac{p_1}{1-p_1}) + \beta_b \mathbb{1}_{(\eta_s>1)} log(\eta_s) + \beta_w \mathbb{1}_{(\eta_s\leq 1)} log(\eta_s), \qquad \text{(B.1)}$$

where p_2 is the elicited beliefs after receiving the ball signal. The coefficient α measures responsiveness to the prior p_1, the posterior beliefs after the color signal, and it is prior in the regression. $\mathbb{1}_{(\cdot)}$ is an indicator function. β_b and β_w separately measure the responsiveness to signals that are indicative of the true state G vs. the true state B. In other words, β_b and β_w separately measure responsiveness to the black ball and white ball. Similarly, the following regression is for TBT sessions:

$$log(\frac{p_2}{1-p_2}) = \alpha log(\frac{p_1}{1-p_1}) + \beta_{bb} \mathbb{1}_{(\eta_s>1)} log(\eta_s) + \beta_{bw} \mathbb{1}_{(\eta_s=1)} log(\eta_s) + \beta_{ww} \mathbb{1}_{(\eta_s<1)} log(\eta_s), \quad \text{(B.2)}$$

[3] See Grether (1980) and Oprea and Yuksel (2020) for example. The regression has been used to study subjects' relative responsiveness to different signals and their prior.

where, β_{bb}, β_{bw}, and β_{ww} estimate traders' responsiveness to two black balls, one black ball, and zero black balls, respetively. Note that regression (B.2) would only estimate β_{bb}, and β_{ww}. This is because the likelihood ratio (η_s) of realizing one black ball and one white ball equals to one due to the symmetry of the balls in the two bags, so the term drops out because $log(\eta_s) = 0$.

If traders react to priors p_1 and ball signals in a perfect Bayesian way, we should have $\alpha = \beta_b = \beta_w = 1$ in equation (B.1) and $\alpha = \beta_{bb} = \beta_{ww} = 1$ in equation (B.2). Also, if traders react to black ball(s) and white ball(s) symmetrically, we should have $\beta_b = \beta_w$ and $\beta_{bb} = \beta_{ww}$.

Table B4 provides the regression results. For both OBT and TBT regressions, the fourth to the last row shows the p-values associated with testing if the responsiveness is the same as perfect Bayesian. Only the green traders in the control under OBT exhibit Bayesian updating since we fail to reject the null hypothesis. Other than that, however, for both OBT and TBT, the estimated coefficients reject perfect Bayesian's responsiveness: traders' (regardless of the colors and the treatments) responsiveness to their prior p_1 and the ball signals are significantly different. This pattern suggests highly non-Bayesian responsiveness to the ball signals.

In the OBT sessions, the joint test between β_b and β_w, on the third to the last row, suggests that traders in the control react to both the black and white ball symmetrically. In the two treatments, however, the joint test suggests that green traders' responsiveness to the black ball, β_b, is significantly greater than their responsiveness to the white ball, β_w; and the responsiveness from the red traders is the opposite: red traders, for the most part, react more to the white ball. This finding indicates that

OBT Sessions						
	Control		Normal		Intense	
VARIABLES	Green	Red	Green	Red	Green	Red
α	0.858	0.404***	0.644***	0.830*	0.801**	0.741***
	(0.167)	(0.0970)	(0.0887)	(0.0997)	(0.0771)	(0.0839)
β_b	0.979	1.046	1.308***	0.850	0.979	0.561***
	(0.0715)	(0.101)	(0.101)	(0.0987)	(0.0771)	(0.0805)
β_w	0.949	0.972	0.563***	0.971	0.711***	1.255**
	(0.0884)	(0.0813)	(0.0978)	(0.122)	(0.103)	(0.102)
$H_0 : \alpha = \beta_b = \beta_w = 1$	0.796	0.000	0.000	0.030	0.036	0.000
$H_0 : \beta_b = \beta_w$	0.792	0.578	0.000	0.461	0.051	0.000
Observations	160	160	152	154	154	159
R-squared	0.687	0.684	0.746	0.661	0.778	0.725
TBT Sessions						
	Control		Normal		Intense	
VARIABLES	Green	Red	Green	Red	Green	Red
α	1.169	0.954	1.067	0.622**	0.619**	0.870
	(0.148)	(0.172)	(0.112)	(0.144)	(0.132)	(0.0982)
β_{bb}	0.717***	0.480***	0.667***	0.762	0.490***	0.570***
	(0.0919)	(0.109)	(0.0884)	(0.226)	(0.0713)	(0.131)
β_{ww}	0.545***	0.536***	0.758*	0.952	0.691***	0.564***
	(0.110)	(0.0949)	(0.120)	(0.163)	(0.0921)	(0.137)
$H_0 : \alpha = \beta_{bb} = \beta_{ww} = 1$	0.000	0.000	0.000	0.011	0.000	0.001
$H_0 : \beta_{bb} = \beta_{ww}$	0.266	0.722	0.618	0.531	0.111	0.976
Observations	62	63	60	59	59	59
R-squared	0.661	0.680	0.805	0.682	0.765	0.699

Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Note: The significance level for each estimated coefficient is associated with testing if the coefficient is significantly different from 1. The data used for the regressions are only elicited beliefs of question 1. Standard errors are clusttered at period level.

Table B4: Estimation of Traders' Responsiveness to Ball Signals

traders have stronger responsiveness to the private signal in favor of their preferred state.

For the normal and the intense treatments in the TBT sessions, by contrast, although the coefficients of β_{bb} and β_{ww} differ economically, the joint tests suggest that such difference is not statistically significant, indicating that a more precise private signals can help traders react to the ball signals more symmetrically. The control in the TBT sessions provides the same symmetric responsiveness to the ball signal.

Motivated traders respond to the ball signals asymmetrically: they react more strongly to the ball in line with their IQ motivations. However, a more precise ball signal largely eliminates such asymmetric responsiveness even if traders hold polarized motivated beliefs.

B.1.21 The Gap Between the CDFs

OBT sessions

	Control	Normal	Intense
Question 1			
Pre-1 (green - red)	0.19	4.40	9.54
Pre-2 (green - red)	0.29	2.50	5.95
Post (green - red)	0.16	2.34	5.30
Question 2			
Pre-1 (green - red)	37.36	43.21	46.11
Pre-2 (green - red)	33.85	40.47	43.12
Post (green - red)	22.14	32.65	35.18
Question 3			
Pre-1 (green - red)	0.0087	0.063	0.0491
Pre-2 (green - red)	0.0069	0.0033	0.0175
Post (green - red)	0.0023	0.0191	0.0114
Shares			
Green - Red	0.028	0.154	0.145
Belief Errors			
(Q1) Pre - post	0.30	0.11	0.10
(Q2) Pre - post	0.28	0.10	0.28
(Q3) Pre - post	0.45	0.48	1.12

TBT sessions

	Control	Normal	Intense
Question 1			
Pre-1 (green - red)	0.087	9.38	6.99
Pre-2 (green - red)	0.18	4.23	2.94
Post (green - red)	0.12	1.84	2.11
Question 2			
Pre-1 (green - red)	45.33	46.69	39.78
Pre-2 (green - red)	39.78	36.67	33.89
Post (green - red)	45.75	30.25	21.85
Question 3			
Pre-1 (green - red)	0.0042	0.0083	0.0161
Pre-2 (green - red)	0.0188	0.0278	0.0381
Post (green - red)	0.0264	0.0835	0.083
Shares			
Green - Red	0.0098	0.063	0.129
Belief Errors			
(Q1) Pre - post	1.46	0.42	2.34
(Q2) Pre - post	1.55	0.38	1.43
(Q3) Pre - post	0.80	4.33	1.21

Note: The gap between two CDFs are measured by the L_2 distance: $\{\int_0^1 (F_a(x) - F_b(x))^2 \, \mathrm{d}x\}^{1/2}$. The left (right) table is for OBT (TBT) sessions. Numbers under questions 1, 2, 3, and Shares provide the distance between green and red traders' elicited beliefs and their final asset holdings. Numbers under Belief Errors are the distance between the pre-trading belief errors and the post-trading belief errors. For question 2, since the dividends are different in the intense treatment compared with the other two treatments, the elicited beliefs are all rescaled to [0,100], which is comparable across the three treatments.

Table B5: The Distance Between two CDFs

OBT sessions

	Control	Normal	Intense
Question 1			
Pre-1 (green - red)	0.002	0.000	0.000
Pre-2 (green - red)	0.050	0.000	0.000
Post (green - red)	0.234	0.000	0.000
Question 2			
Pre-1 (green - red)	0.053	0.000	0.000
Pre-2 (green - red)	0.089	0.000	0.000
Post (green - red)	0.400	0.033	0.000
Question 3			
Pre-1 (green - red)	0.410	0.001	0.004
Pre-2 (green - red)	0.296	0.796	0.138
Post (green - red)	0.721	0.143	0.852
Shares			
Green - Red	0.071	0.000	0.000
Belief Errors			
(Q1) Pre - post	0.313	0.275	0.103
(Q2) Pre - post	0.174	0.179	0.006
(Q3) Pre - post	0.765	0.374	0.148

TBT sessions

	Control	Normal	Intense
Question 1			
Pre-1 (green - red)	0.569	0.000	0.000
Pre-2 (green - red)	0.790	0.000	0.000
Post (green - red)	0.931	0.057	0.021
Question 2			
Pre-1 (green - red)	0.287	0.000	0.000
Pre-2 (green - red)	0.420	0.002	0.003
Post (green - red)	0.769	0.055	0.186
Question 3			
Pre-1 (green - red)	0.727	0.744	0.454
Pre-2 (green - red)	0.670	0.221	0.640
Post (green - red)	0.911	0.046	0.077
Shares			
Green - Red	0.957	0.122	0.015
Belief Errors			
(Q1) Pre - post	0.005	0.435	0.000
(Q2) Pre - post	0.000	0.958	0.000
(Q3) Pre - post	0.336	0.015	0.254

Note: The numbers are the p-value for the WSR hypothesis tests. Questions 1, 2, 3, and Shares test the difference between green and red traders' elicited beliefs and final asset holdings (Pre-1, Pre-2, Post refer to pre-trading survey 1, pre-trading survey 2, and post-trading survey, respectively). Belief Errors test the difference between pre-trading and post-trading belief errors (Q1, Q2, and Q3 refer to question 1, question 2, and question 3, respectively).

Table B6: Wilcoxon Signed-Rank(WSR) Test of the CDFs

	OBT Sessions				TBT Sessions			
	Green		Red		Green		Red	
	$\|p_2 - p_1\|$	$\|p_3 - p_2\|$	$\|p_2 - p_1\|$	$\|p_3 - p_2\|$	$\|p_2 - p_1\|$	$\|p_3 - p_2\|$	$\|p_2 - p_1\|$	$\|p_3 - p_2\|$
Question 1								
C vs N	0.000	0.000	0.058	0.057	0.714	0.354	0.568	0.142
C vs I	0.000	0.004	0.012	0.215	0.941	0.739	0.794	0.789
N vs I	0.654	0.373	0.487	0.381	0.510	0.238	0.794	0.094
Question 2								
C vs N	0.005	0.003	0.053	0.012	0.341	0.281	0.398	0.091
C vs I	0.000	0.000	0.014	0.000	0.780	0.050	0.667	0.427
N vs I	0.558	0.665	0.769	0.128	0.319	0.428	0.276	0.110

Note: Wilcoxon Signed-Rank (WSR) Test for the update magnitudes in figure B13 and figure B14. The numbers in the table are the p-values of the hypothesis tests results. Column titles specify which gap this result measures under which color (green or red) and which session (OBT or TBT). Row names indicate which two CDFs are being tested: "C vs N", "C vs I", "N vs I" represent "control vs normal", "control vs intense", and "normal vs intense", respectively.

Table B7: Wilcoxon Signed-Rank Test for the Update Magnitudes

B.1.22 Alternative Measures of Information Aggregation

This section adapts two models from Page and Siemroth (2021) to study market efficiency. All the observed prices under this subsection are rescaled to the interval of [0,1] (dividend of the bad/good state is set to equal to 0/1) so that the model can be applied to all treatments consistently.

After rescaling the prices into the interval [0,1], the RE-price is simply the Bayesian posterior (p^*) of state G conditional on all the available signals (denoted as N) in the market. The first model assumes the market can incorporate a subset of the available signals into the prices. The observed prices are set as if the market incorporates a random sample (without replacement) of signals $n = \lceil \lambda N \rceil$, where $\lceil x \rceil$ denotes the least integer greater than or equal to x. The parameter of interest is $\lambda \in [0, 1]$, the proportion of all available signals used. The model uses the maximum likelihood method to estimate the best fitted λ given the most relevant data, the last transaction price (the average of the last two prices is also used for the estimation). The higher the λ, the market is more efficient since it incorporates more private information into the prices. For instance, the market is fully efficient ($\lambda = 1$) if it incorporates all the available signals N into the prices. For detailed math derivations and the complete explanation of the logic of the model, please see the original paper (Page and Siemroth (2021)).

Table B8 reports the estimation results. The estimated λ might not be unique. For example, when $\lambda = 0.10, 0.11, \text{or } 0.12$, $\lceil \lambda \times 16 \rceil$ have the same value, which is 2. In this case, only the largest λ is reported in table B8. Let's first focus on the first two columns where the estimation calculated by using only the last price. In the OBT

	Last Price		Last Two Prices	
	OBT	TBT	OBT	TBT
Data	λ	λ	λ	λ
Control	0	0.12	0	0.12
	[0.00,0.12]	[0.00,0.44]	[0.00,0.12]	[0.00,0.34]
Normal	0	0.18	0	0.12
	[0.00,0.25]	[0.00,0.50]	[0.00,0.25]	[0.00,0.47]
Intense	0.12	0.18	0.12	0.18
	[0.00,0.25]	[0.06,0.43]	[0.00,0.25]	[0.06,0.43]

Note: Recall that there are 8 trading periods in each session and 5(2) sessions for each treatment in OBT(TBT); therefore, there are $8 \times 5 = 40(8 \times 2 = 16)$ observations in OBT(TBT). The 95% confidence intervals below the estimated λ are calculated via the non-parametric percentile bootstrap method, which resamples (with replacement) the market periods within each treatment 100 times to determine the final confidence intervals.

Table B8: Proportion of Reflected Information Measured by λ

sessions, the market prices do not incorporate any information from the market in the control and normal treatment since $\lambda = 0$ in these two cases. This finding is consistent with the graphical evidence of transaction prices that show most of the transaction prices are around the mid-point of the dividends. However, increasing the financial stakes helps improve the information aggregation since λ in the intense treatment equals 0.12. From the TBT sessions, we can see that all the λ across different treatments are increased, indicating that increasing the precision of the second private signal improves the market efficiency. But the estimated λ remain similar across three treatments. We observed similar results in section 2.3.3. In addition, the results on the last two columns in Table B8 remain similar if we use the average of the last two prices.

The second model adapted from Page and Siemroth (2021) is the "price accuracy" measured by ψ. As shown in Equation (B.3), it measures the distance between the observed price p_m and the prior (no) information price 0.5, relative to the distance between the RE-price and the prior information price in a specific market m:

$$\psi_m = min\left\{1, \frac{p_m - 0.5}{\text{RE-price}_m - 0.5}\right\}, \tag{B.3}$$

where, the minimum operator ensures that ψ cannot be larger than 1 when the observed price p_m overreacts to the information. Therefore, ψ takes a value between $[0, 1]$, and a larger ψ indicates a higher price accuracy given an observed price p_m. Similar to the model that estimates λ, the estimation of ψ is also based on the last transaction price as well as the average of the last two prices in each period. To estimate ψ, the model runs an ordinary least squares (OLS) regression as follows to find out the average price accuracy ψ within a treatment:

$$Y_m = \psi(\text{ RE-price}_m - 0.5) + \epsilon_m$$

with,

$$Y_m = \begin{cases} min\{p_m - 0.5,\ \text{RE-price}_m - 0.5\}, \text{if RE-price}_m > 0.5, \\ max\{p_m - 0.5,\ \text{RE-price}_m - 0.5\}, \text{if RE-price}_m < 0.5. \end{cases}$$

	Last Price					
	OBT			TBT		
	Control	Normal	Intense	Control	Normal	Intense
ψ	0.01	0.09	0.22**	0.38*	0.30	0.40*
	(0.11)	(0.08)	(0.08)	(0.15)	(0.20)	(0.14)
R^2	0.00	0.03	0.18	0.32	0.13	0.36
Obs.	40	40	40	16	16	16
	Last Two Prices					
	OBT			TBT		
	Control	Normal	Intense	Control	Normal	Intense
x	−0.03	0.06	0.22**	0.36*	0.27	0.36*
	(0.10)	(0.07)	(0.07)	(0.15)	(0.20)	(0.15)
R^2	0.00	0.02	0.20	0.28	0.11	0.28
Obs.	40	40	40	16	16	16

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

Note: The table displays the OLS estimates for the price accuracy ψ. Five sessions in each treatment in OBT provide 40 observations, and two sessions in each treatment in TBT provide 16 observations.

Table B9: Price Accuracy Measured by ψ

Table B9 displays the estimation results, which is consistent with the estimation of λ. Using both last price and the average of the last two prices, in OBT sessions, both the control and the normal sessions have a low ψ that does not significantly differ from zero, but the intense treatment has a $\psi = 0.22$. In TBT sessions, the estimated ψ remain similar across three treatments.

B.1.23 Beliefs at Key Junctures (Subject Average Measure)

This section replicates the analysis of section 2.3.2 but using subject average measure. Each observation takes the average across different traders in each session conditional on his/her color. Figure B27, B28, and B29 are traders' belief updating processes under OBT sessions. Figure B30, B31, and B32 are traders' belief updating processes under TBT sessions. From the graph, we can see that the patterns are largely the same as the patterns in the main text. Therefore, the results remain the same.

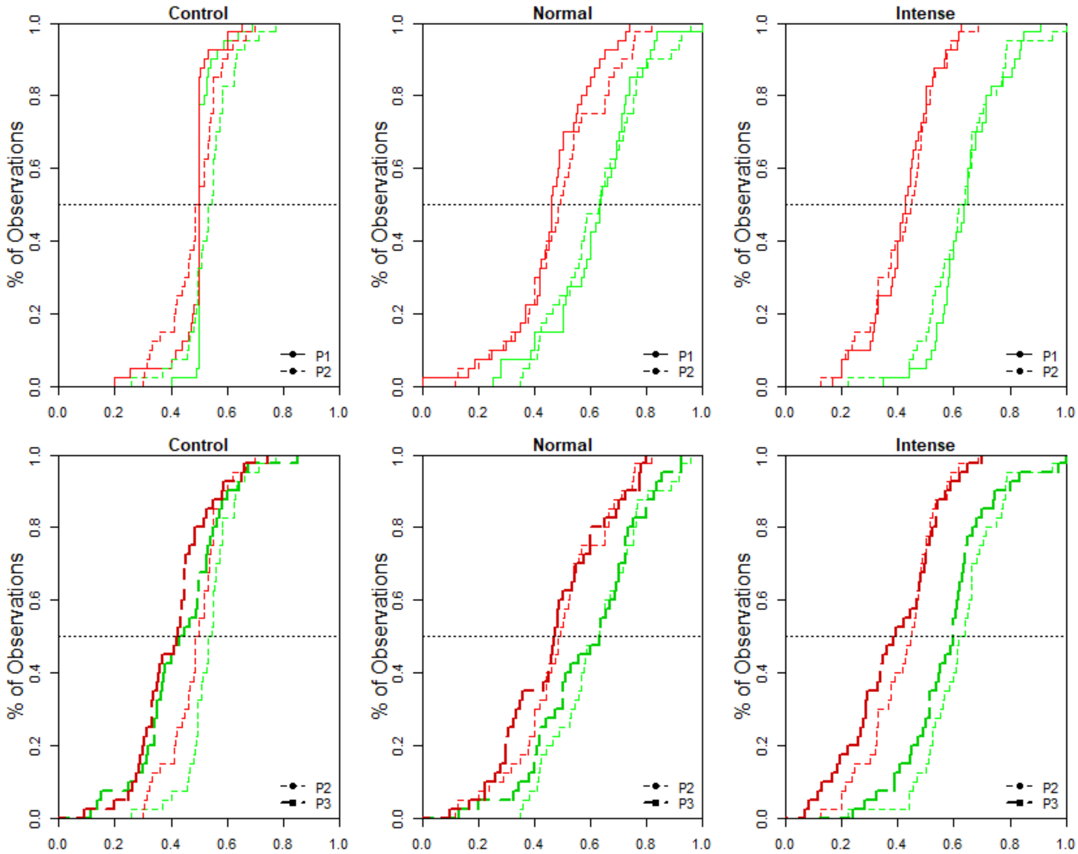


Figure B27: Belief Updating Question 1 (OBT)

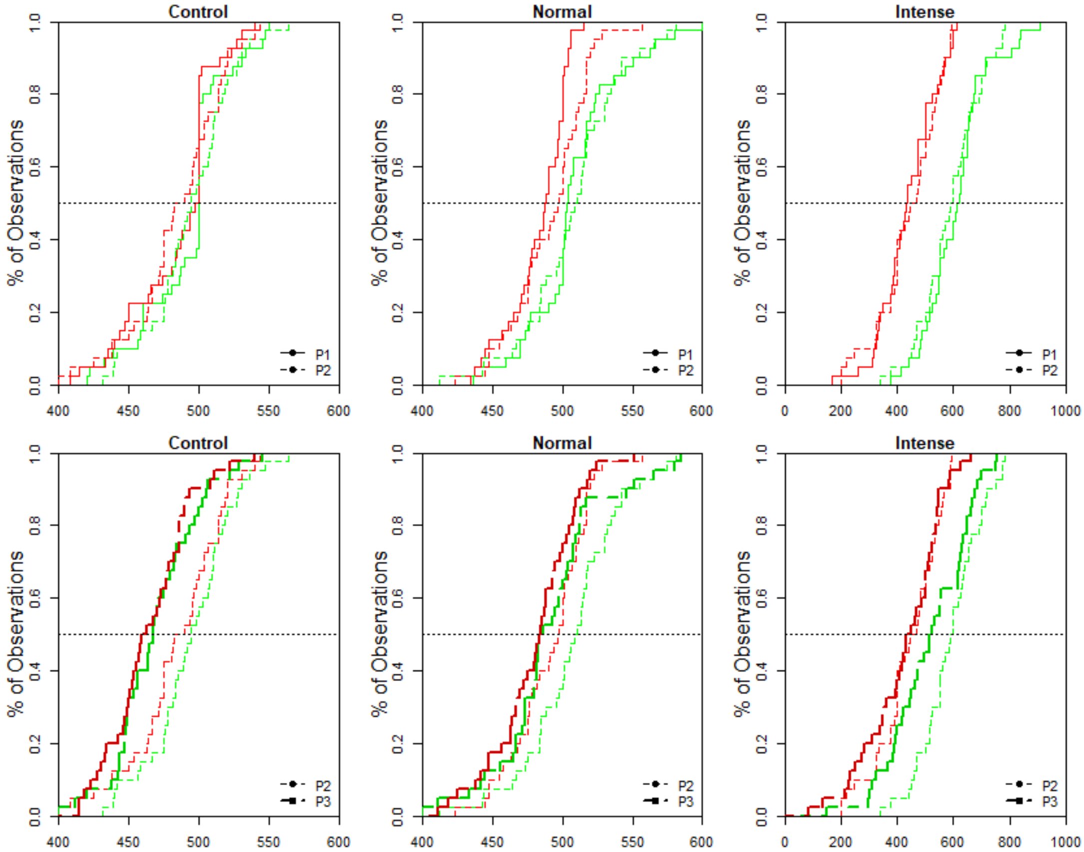


Figure B28: Belief Updating Question 2 (OBT)

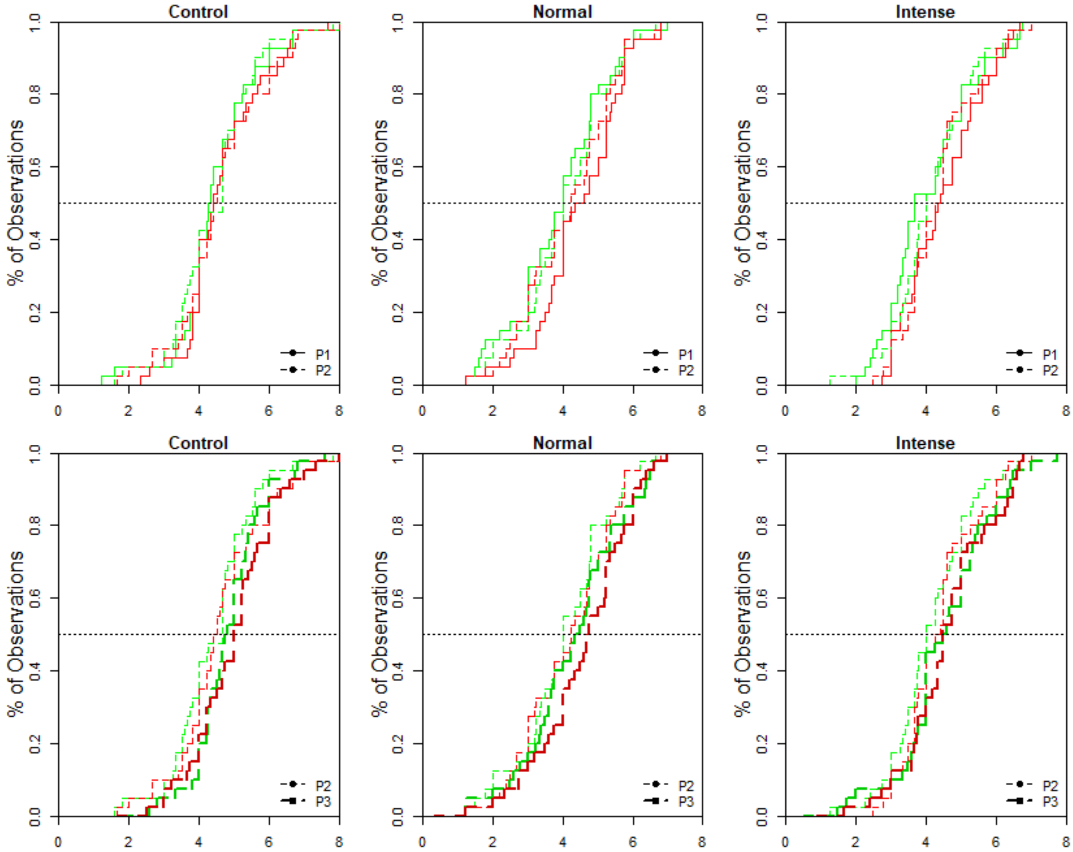


Figure B29: Belief Updating Question 3 (OBT)

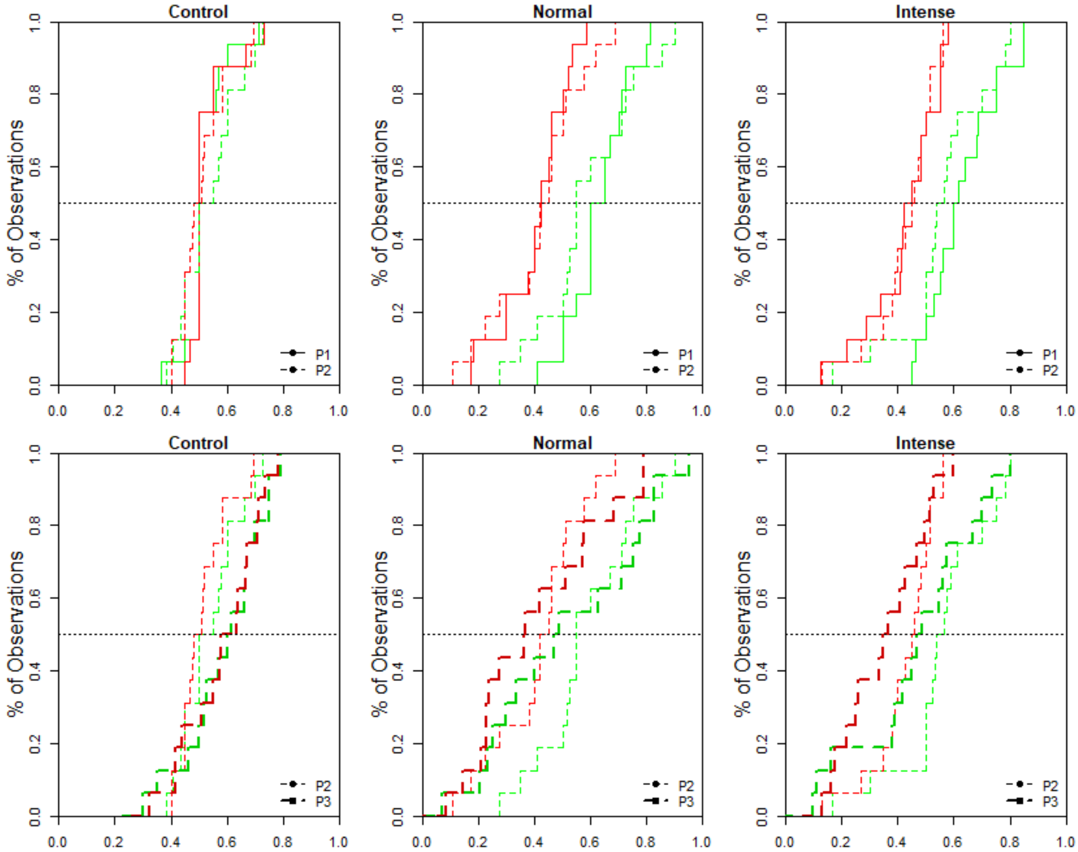


Figure B30: Belief Updating Question 1 (TBT)

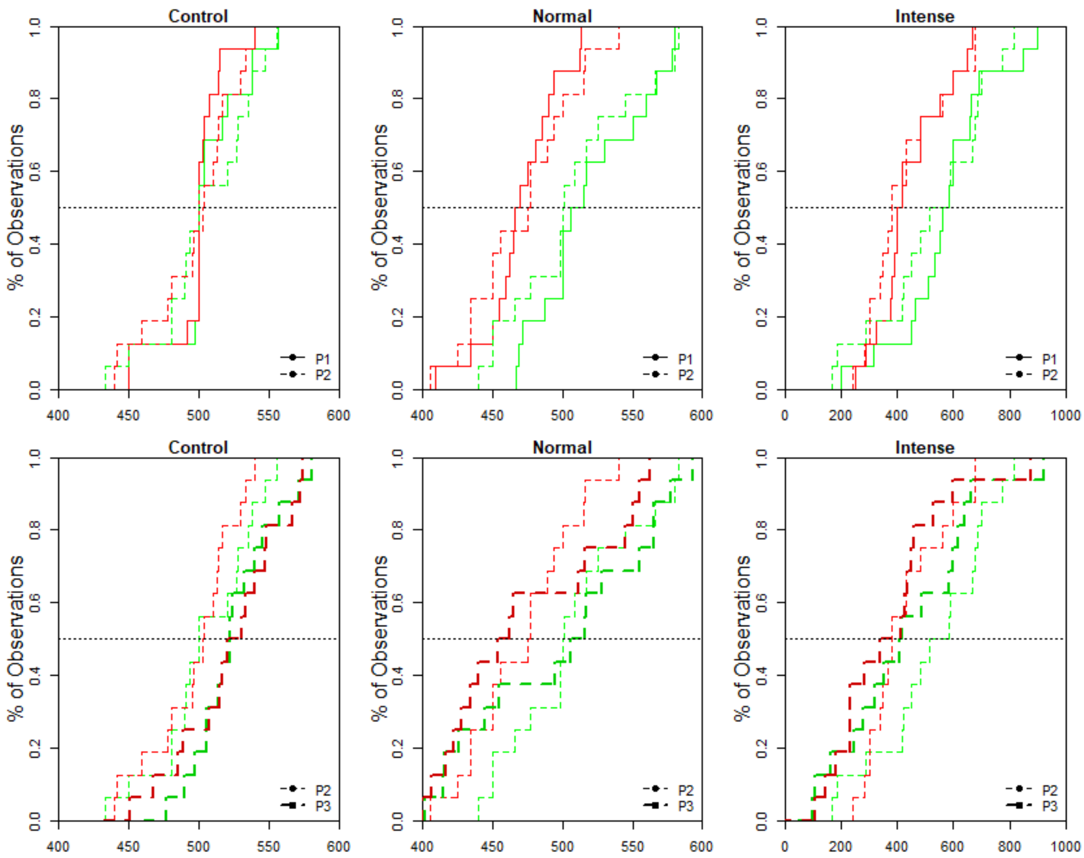


Figure B31: Belief Updating Question 2 (TBT)

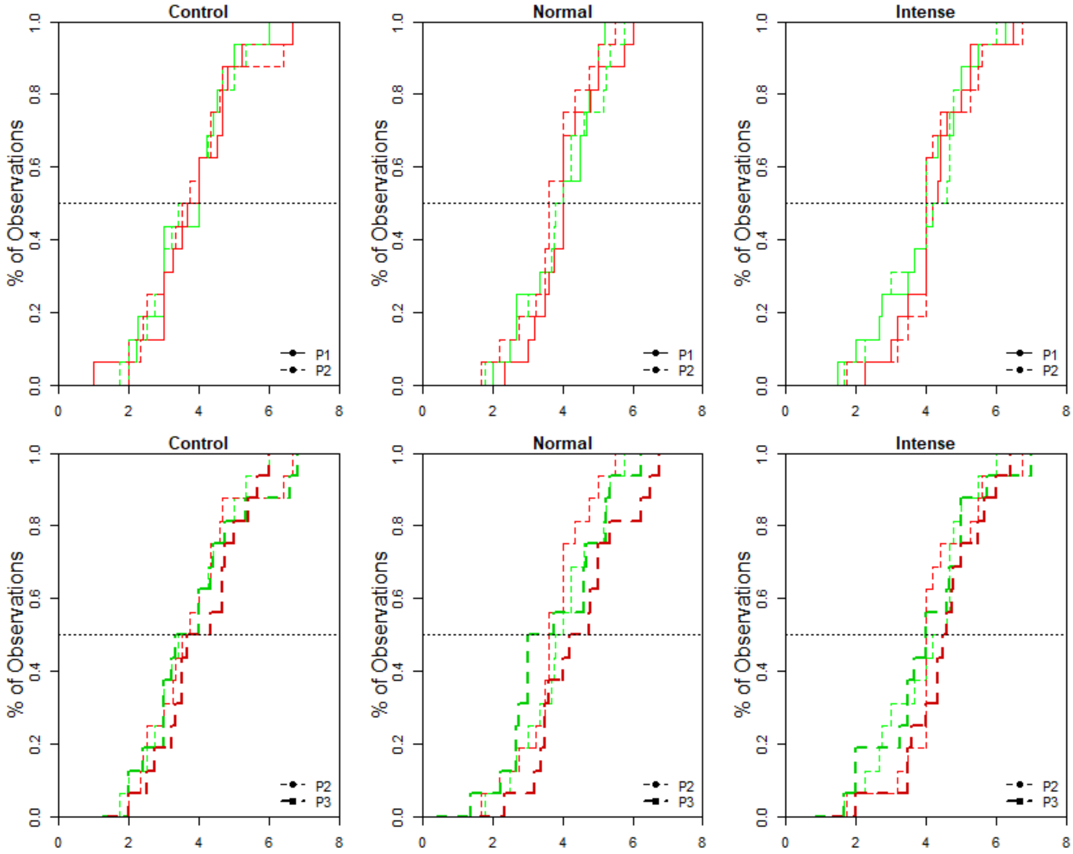


Figure B32: Belief Updating Question 3 (TBT)

B.1.24 Final Asset Allocations (Subject Average Measure)

This section replicates the analysis of Figure 2.7 but using the subject average measure. Each observation takes the average across different traders in each session, conditional on their color. Figure B33 and Figure B34 are under OBT and TBT sessions, respectively. The graph shows that traders' final asset holdings are similar between green and red traders in control. But the share holdings are diverged between green and red traders in the two treatments. The result is, therefore, the same as result 2.3.3.2.

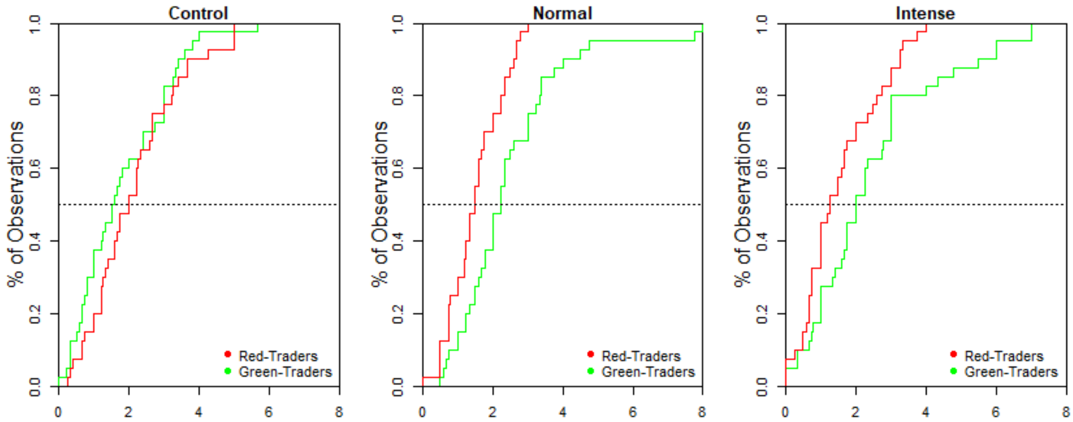


Figure B33: Share Holdings (OBT)

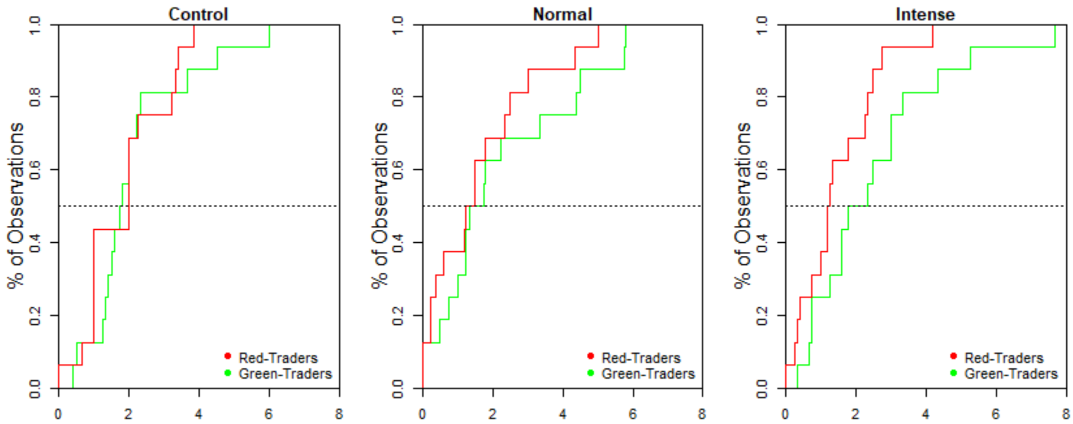


Figure B34: Share Holdings (TBT)

B.1.25 Belief Errors (Subject Average Measure)

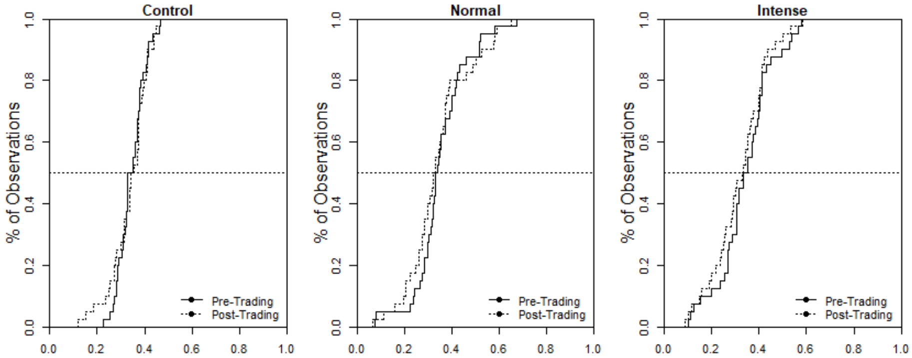


Figure B35: Belief Errors Question 1 (OBT)

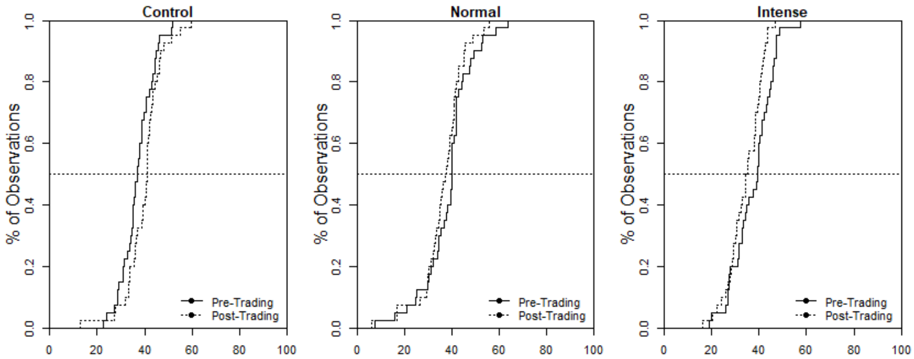


Figure B36: Belief Errors Question 2 (OBT)

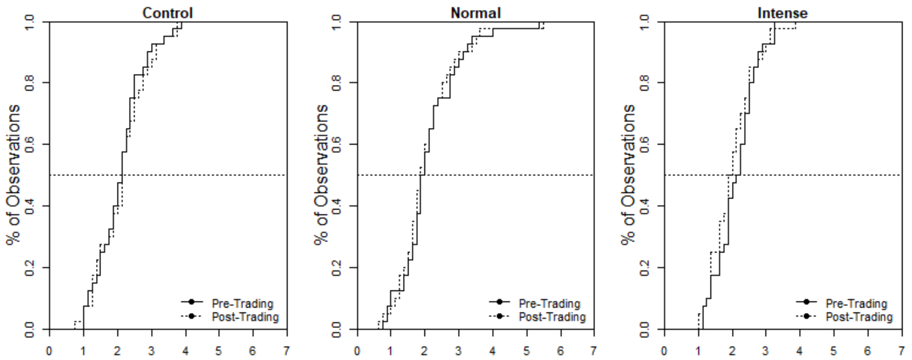


Figure B37: Belief Errors Question 3 (OBT)

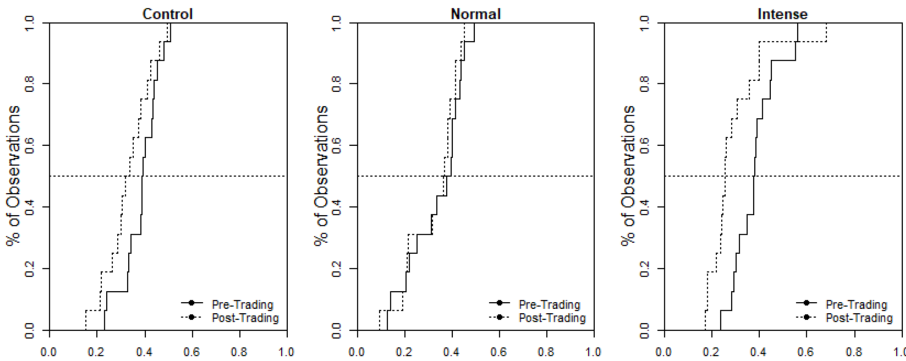


Figure B38: Belief Errors Question 1 (TBT)

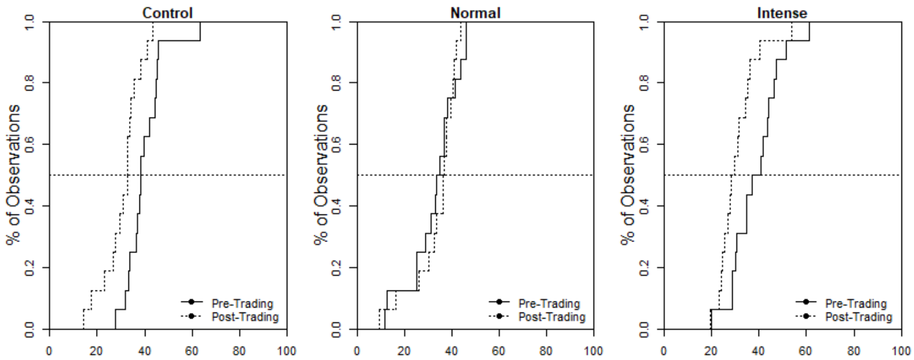


Figure B39: Belief Errors Question 2 (TBT)

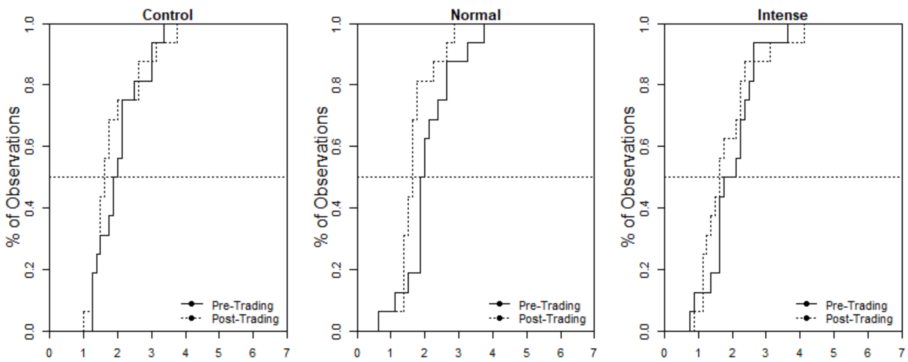
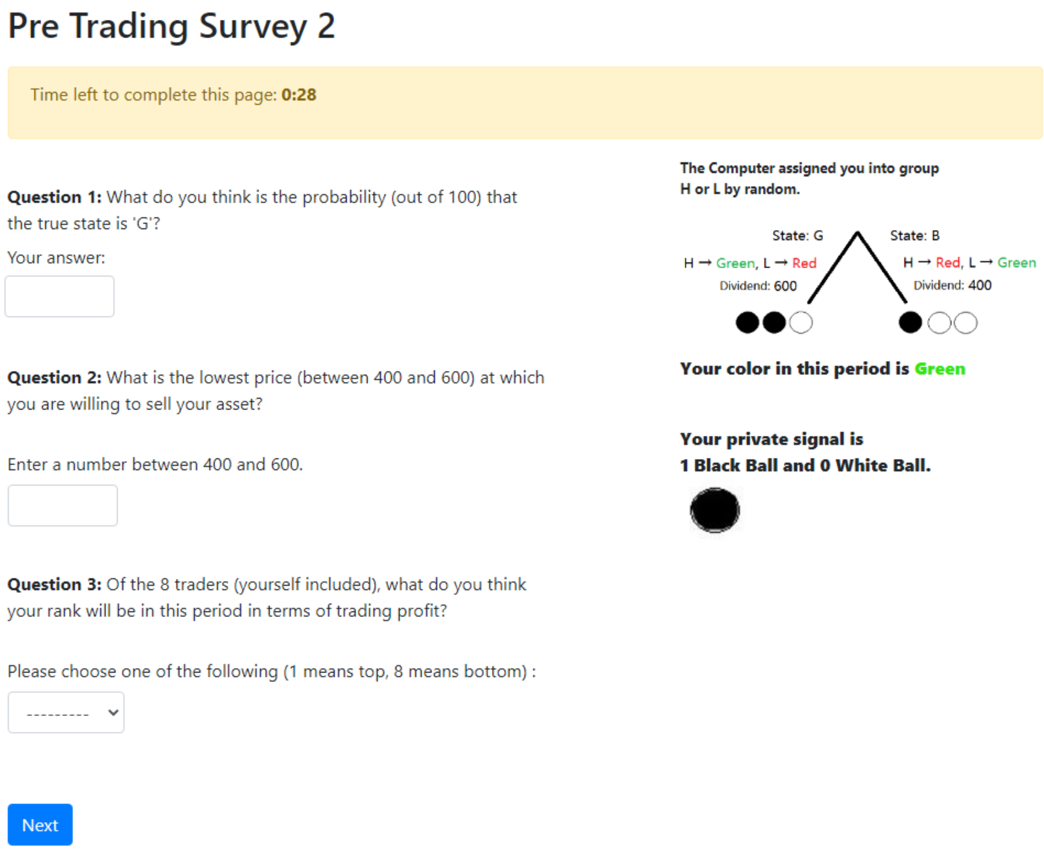


Figure B40: Belief Errors Question 3 (TBT)

B.2 Experiment Instructions and User Interface

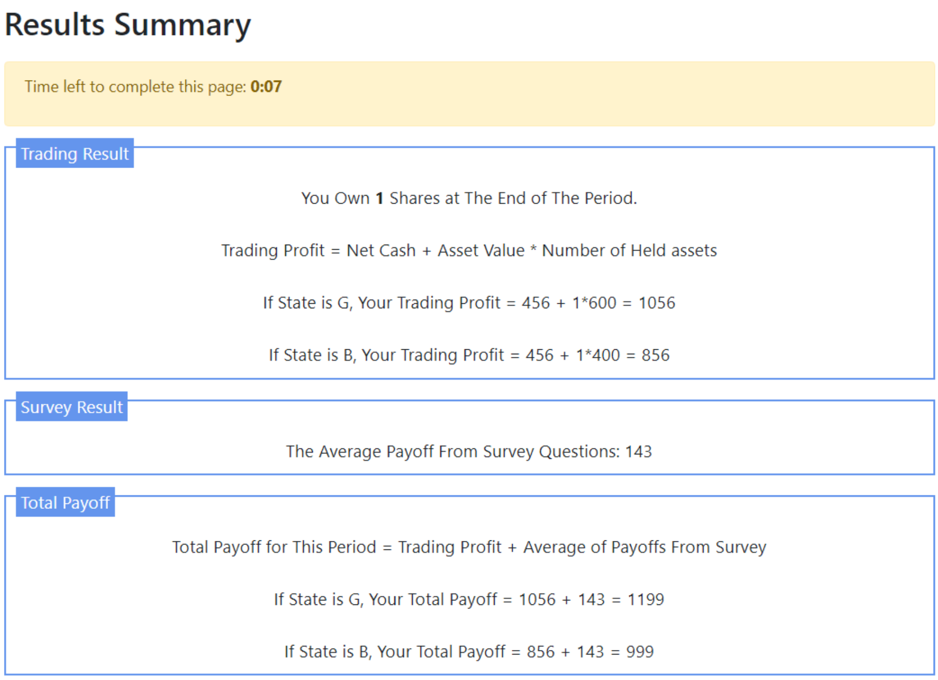
B.2.1 User Interface



Note: If the session is control, then the statement on the top of the private signals says, "The computer assigned you into group H or L by random"; if the session is one of the treatments, the statement would say "The computer assigned you into group H(L) if your IQ rank is higher(lower) than your paired player.

Figure B41: UI of Pre-Trading Survey 2

Figure B41 shows an example of Pre-trading Survey 2 that asks the three questions after the second private signal was revealed. Please note that the only difference between Pre-trading Survey 2 and Pre-trading Survey 1 is that the latter only reveals the first color signal. Pre-trading survey 1, pre-trading survey 2, and post-trading survey all ask the same three questions. Question 1 asks What do you think is the probability (out of 100) that the true state is 'G'? Question 2 asks What is the lowest price (between 400 and 600)[4] at which you are willing to sell your asset? Question 3 asks Of the 8 traders (yourself included), what do you think your rank will be in this period in terms of trading profit? All elicitation tasks are incentive-compatible, so the best strategy for each trader is to report their beliefs truthfully.



Note: The other feedback pages are "Trading Result," "Survey Result," and "Total Payoff," respectively. This figure is the UI of the "Results Summary" page, which is the last feedback page that summarizes all the previous feedback pages.

Figure B42: UI of Performance Summary

[4] If the treatment is intense, the question would ask subjects to enter a number between 0 and 1000.

B.2.2 Experiment Instructions

B.2.2.1 Instructions for Normal Treatment under One Ball Treatment

INSTRUCTIONS

Welcome! This is an experiment in the economics of decision-making. If you pay close attention to the instructions and make good decisions, you can earn a significant sum of money which will be paid in cash at the end of the session.

Please turn off your cell phone and other communications devices now, and do not communicate with other participants. Please feel free to direct your questions to the experimenter. We expect and appreciate your cooperation.

The Basic Idea.

You and the other 7 traders will trade an asset using computerized trading screens. The asset has two possible values. If the state is good (G) then each unit of the asset pays 600 points to the trader who holds it at the end of the trading period, but it pays only 400 points if the state is bad (B). Before each trading period, the computer tosses a fair coin, and assigns state G to all units of the asset if the coin comes up heads or B if tails. That is, at the beginning of each period each state is equally likely, and the same state applies to all traders.

You will not be told the true state but, as explained below, you will get hints from the computer in the form of private signals. After receiving private signals, you will be **asked how likely** you now think that the true state is G. Then you will be able to **trade with other participants**. Since they also get private signals that may differ from yours, the trading prices you see may also provide hints about the true state and the asset value. After the trading period is over you will be asked about your **final belief** about the true state in this period, based on both your private signals and on other traders' offers and trades. Finally, at the end of each period, you will receive feedback on (a) your trading profits for both possible states, (b) your payoffs from answering questions about your beliefs, and (c) your total payoffs for the period, which equals to (a) + (b).

Each period lasts approximately 3 minutes. There will be between 7 and 12 trading periods in today's experiment. The total of all points you earn in all periods will be converted into US dollars and paid to you in cash.

Test Scores

The 10 questions you just completed before this instruction are often used as non-verbal tests of intelligence. Scores on these IQ questions are ranked among today's participants from highest (top, 1st) to lowest (bottom, 8th). Tie scores are broken randomly.

Two Private Signals At the beginning of each period, the computer first randomly sorts all 8 traders into 4 pairs (2 traders each pair). In each pair, the trader with higher (resp. lower) rank on the ten IQ questions is assigned to the H (resp. L) group. Then the computer will randomly determine the true state. As shown in Figure 1, when the state is G, the H-group traders will be marked as Green, and L-group traders will be marked as Red; the marking will be reversed if the state is B.

[**First Private Signal: color**] You will NOT know whether you were assigned to the H or L group. But you will be told your group's color after the true state is determined. Your color is your first private signal. It may help you guess how likely the true state is G.

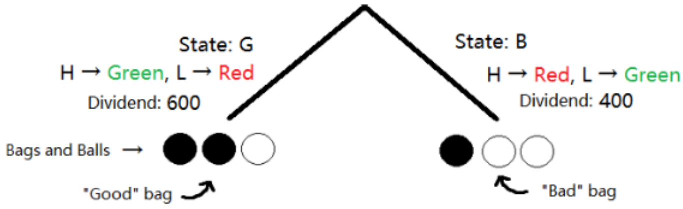


Figure 1

[**Second Private Signal: ball**] If the true state is G then the computer draws a ball **randomly** from the "Good bag" that contains 2 black balls and 1 white ball. Similarly, if the true state is "B", the computer will draw a ball **randomly** from the "Bad bag" that contains only 1 black ball plus 2 white balls. The computer will show you the drawn ball. This is your second private signal. If your ball is black, then state G will be more likely because a black ball is more likely drawn from the "good bag", which contains more black balls. Likewise, if your ball is white then it is more likely to have come from the "bad bag", so state B is somewhat more likely.

Surveys

Each period you will be surveyed three times. Each survey has the same three questions. The more accurate your answers, the more extra payment you will receive. The questions are as follows.

Question 1: What do you think is the probability (out of 100) that the true state is 'G'? Please enter a number that reflects your belief; if you think that, given your current private signal(s), it is 65% likely the true state is Good, then type 65 into the answer box.

Question 2: What is the lowest price (between 400 and 600) at which you are willing to sell your asset? You will soon have the opportunity to sell the asset

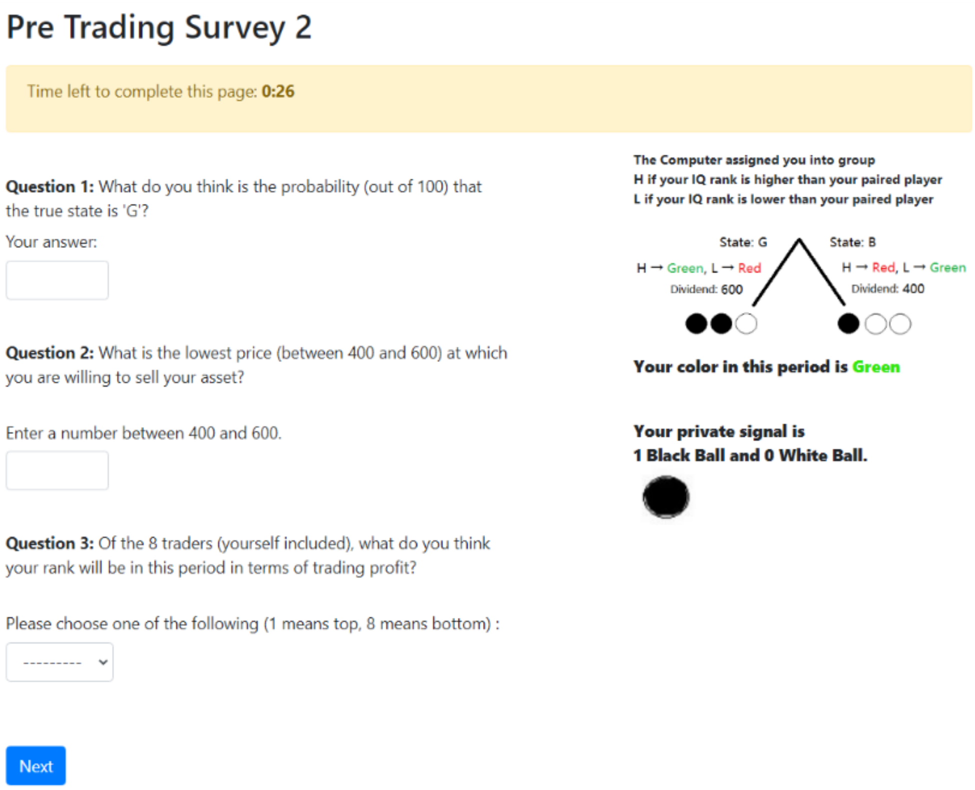


Figure 2

and must decide on what price to accept for each unit, given your beliefs about whether it will eventually pay 400 (if B) or 600 (if G). Please type your lowest acceptable price into the answer box for this question.

Question 3: Of the 8 traders (yourself included), what do you think your rank will be in this period in terms of trading profit? (1 means top, 8 means bottom) The computer will rank all traders' payoffs earned in the current trading from the highest (1st) to the lowest (8th). Please guess how your own trading profits will compare to other participants'. For example, if you guess that your trading profits will be the 2nd highest among the 8 traders, then select the number 2 and click "Next".

The first survey comes right after the computer shows you your color, red or green (first private signal). You answer the three questions, click the "Next" button, and then the computer shows you black or white drawn ball (second private signal). Using both private signals to update your beliefs, you answer the three questions again according to your updated beliefs. Figure 2 is an example screenshot for this second survey.

Before taking the third survey, you will have the opportunity to trade the asset in a market. As explained in the ⟨Survey Question Procedure⟩, the payments you receive for answering the survey questions are made in such a way that it is in your best interest to think carefully and respond truthfully to each question.

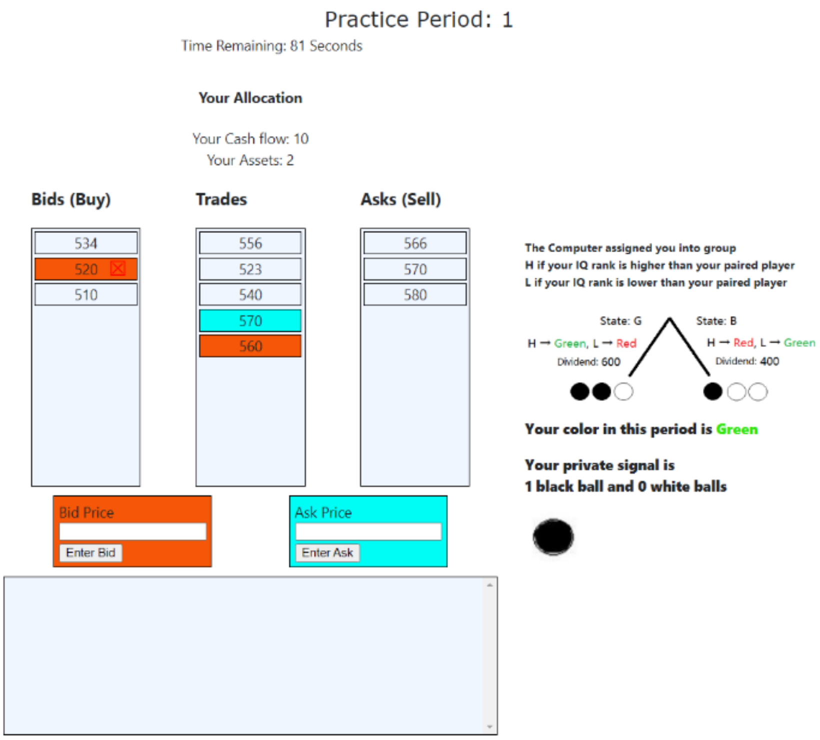


Figure 3

How to trade

Figure 3 shows the computer screen for trading. The right-hand side reminds you of your two private signals. There is a timer on the top counting down the time remaining.

"Your Allocation" box tells you your current cash holding and number of asset units; it will update whenever you trade. You will start with two units of assets at the beginning of each trading period.

"Bids(Buy)" box shows the prices at which traders are offering to buy units of the asset and the "Asks(Sell)" box shows the prices at which sellers are offering to sell units. Bids are ordered from high to low since the highest bids are the most attractive offers to other traders. Likewise, asks are ordered from low to high since the lowest is the most attractive offer to other traders. If two bids/asks have the same price, the earlier offers will be displayed above the later ones. A bid/ask colored in orange/blue indicates your own offer.

"Trades" box in the middle shows the trading history in the market with the most recent trades on the top. Your own purchases/sales are colored in orange/blue.

You submit a bid (buy offer) by typing in your price in the "Bid Price" box and clicking the Enter Bid button. Likewise, to offer to sell a unit, you type in the price in the "Ask Price" box and click Enter Ask. You can place at most 1 bid and 1 ask at the same time. If you enter a new bid or ask, it will replace your old offer. If your bid is higher than the best ask, then you will immediately buy a unit at that best ask. Similarly, entering an ask below the best bid held by someone else amounts to accepting that best bid. You can also directly accept someone else's bid or ask by double clicking on it, and then clicking Accept in the pop-up window. You can buy at most 8 shares of assets.

It is more profitable to buy low and sell high. Check the trading window and see whether your Orange buy prices seem low, and your Blue sell prices seem high compared to the other prices (in white) this period.

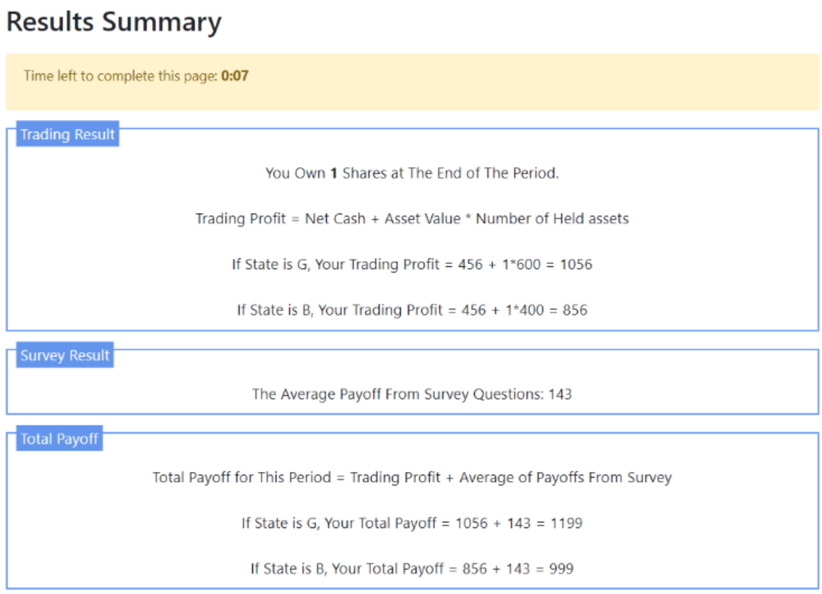
The large "Error Message" box at the bottom of the trading screen will tell you if you made some kinds of mistakes, such as trying to sell when you currently hold zero units of the asset.

How do you get paid for trading in the market? The payoff from trading in each period will be net cash flow (i.e., receipts from asset units sold minus cost of units bought), plus the payments on the assets you hold. For example, if you sold one unit for 580 and then you bought a unit for 560, your net cash flow would be 580 - 560 = 20. Moreover, since you started with 2 units of the asset, sold one unit and then bought one unit, your final asset holding is 2 units. Then your final payoff from trading in this period equals to 20+2*600=1220 if the state turns out to be G, and 20+2*400=820 if the state is B.

Post-Trading Survey

After each trading period, there will be another survey page that asks you the same three questions as in the pre-trading survey about your beliefs. The only difference between the post- and pre-trading survey is that the post-trading survey reminds you of the trading history from the current period, and that history might tell you something about what other traders' private signals are. So taking that history as well as your own private signals into account, you should update your beliefs again before you answer the questions.

In addition, the trading history in this post-trading survey is ordered from high to low. The computer did this descending sorting for you to help you learn your trading strategy. If your buying prices (orange) are, on average, lower than your selling prices (blue), then it means you are making profit. If it is the other way around, you need to adjust your trading strategy.



Feedback Page

After the post-trading survey page, you will see feedback pages similar to that shown below. These pages remind you of your final asset holding. They also show you the average points you earned from answering survey questions and the payoff you earned from trading. They also show the total payoff for this period, which is ”payoff from trading” + ”average payoff from survey questions” (All payoffs are conditional on different states). You will NOT be informed about which state is the true state. After this feedback page, all traders go on together to the next period.

You will play 7 to 12 trading periods. Your final payoff for participating in today’s session equals the sum of the total payoffs from all periods. Your final payoff is converted from experimental points to cash (US Dollars) at the following rate: 1,000 experimental points = $2 in cash.

B.2.2.2 Instructions for Other Treatments

All the instructions use the same format from the previous subsection and they only differ in the following points:

- The instructions for the intense treatment under OBT only changes the dividends to 1000 for Good state and 0 for Bad state.
- The instructions for the control under OBT only changes the "grouping assignment" to state that whether the trader is assigned in the H or L group is completely random.
- The instructions for Control, normal, and intense treatments under TBT sessions only changes the second private signal part and mention that "the computer randomly draws two balls with replacement out of the bag that corresponds to the true state". Graphs in the instructions need to be changed accordingly to show two drawn balls, all else stays the same.

B.2.3 Survey Questions Procedure (Control and Normal)

Survey Questions Procedure

Question 1 (in both the pre and post trading survey): "What do you think is the probability (out of 100) that the true state is 'G'?"

How you will get paid: When you answer this question, you will be offered one unit of the asset. You can earn the dividend from this asset or exchange it with another asset that will be described below.

- Enter what you believe the probability is that the true state of the trading asset is "G". Let us denote this probability as R.
- We will offer you another asset called the N-asset, which is worth 600 with probability P_N and 400 with probability (1-P_N). (The probability P_N is determined randomly between 0 to 99.)
- If $R > P_N$, you keep your trading asset and earn the dividend from the asset. (600 for G and 400 for B).
- If $R \leq P_N$, you exchange your trading asset with us, thus hold the N-asset, and earn the dividend from the N-asset.

Question 2 (in both the pre and post trading survey): "What is the lowest price (between 400 and 600) at which you are willing to sell your asset."

How you will get paid: When you answer this question, you will be offered one unit of the asset. You can either earn the dividend from this asset or sell it back to us.

- You decide on the lowest price at which you are willing to sell your asset and enter that price into the computer.

- We will offer you a price to buy the asset back from you. The price is a random number between 400 and 600.
- If the price you are willing to sell is higher than our purchase price, you will keep your asset and earn that asset's value.
- If the price you are willing to sell is lower than our purchase price, you will sell your asset to us at the price we offered.

Question 3 (in both pre post trading surveys): "Of the 8 traders (yourself included), what do you think your rank will be in this period in terms of trading profit? (1 means top, 8 means bottom)"

How you will get paid: Your payoff from answering this question is determined by the formula $100-(C-R)^2$. Where, C refers to the correct ranking of your payoff among all traders. R refers to your guess, which you entered into the computer.

Your best strategy for answering the three questions is to truthfully report!

Here is why (we use question 2 as an example to explain):

▷ You think this asset is worth at least 500 to you, so 500 is the lowest price you would be willing to sell. This is your true belief. But you enter 550 into the box. You overstate your true belief.

▷ Then the computer will offer a price to buy, and that price is a random number between 400 and 600. Suppose the computer randomly chooses 540 to offer you for the asset.

▷ The computer cannot purchase the asset from you because 540 is lower than the lowest price you would accept to sell (550). So you keep your asset.

▷ This means you only earn 500; however, if you truthfully report 500 into the computer, then the computer would purchase that asset from you at the price of 540, so you would earn 540.

▹ Your earning is higher when you truthfully report.

The payment methods for question 1 and question 3 are designed in the same way, which means there is no circumstance in which offering a probability/price/ranking not equal to your true belief is to your advantage; it can only decrease your earnings.

After you finish reading the written instruction, we will show you a video to explain, visually, what this game is about.

B.2.4 Survey Questions Procedure (Intense)

The "Survey Questions Procedure" of intense treatment only changes the dividends from (400,600) to (0,1000) accordingly, all else stays the same.

B.2.5 Intuitive Example for Truth-telling Mechanism

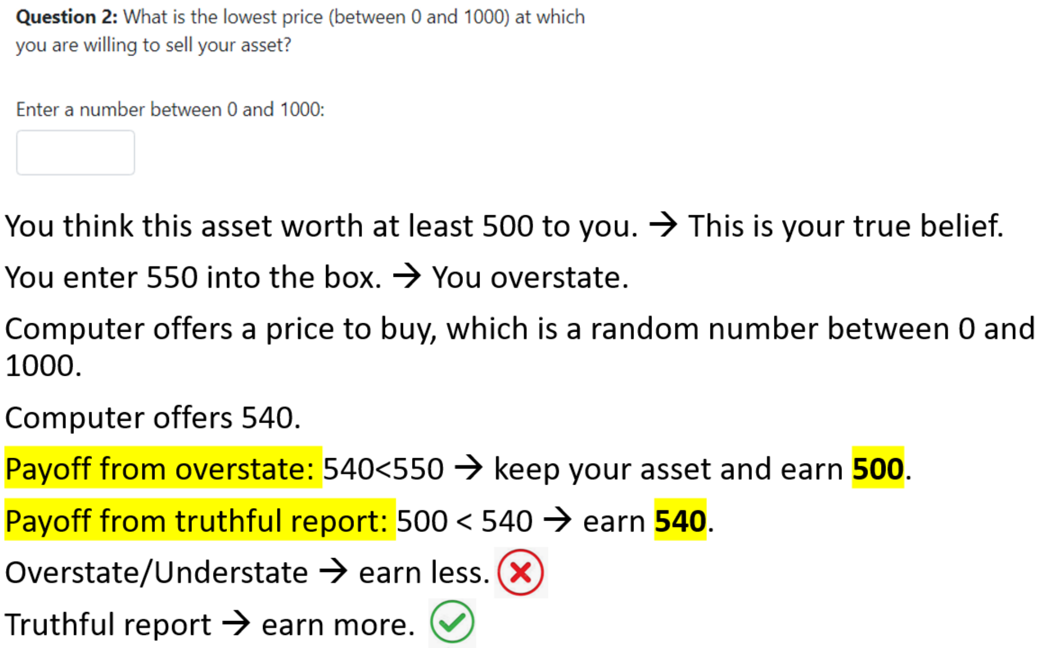


Figure B43: Example Explains Mechanism

Appendix C

Chapter 3 Appendix

C.1 Supplementary Figures and Tables

C.1.1 Notes for Data Cleaning

- Some "miss-reported" data is corrected. One miss-reporting is no answer collected from some survey questions. Subjects might spent too much time thinking of their responses and didn't realize the time was out for the current page. Another miss-reporting is the answers are out of the reasonable range (e.g., the subject accidentally reported a probability of 120% for question 1). To deal with the "miss-reported" data, we used the midpoint of the ranges as their answers for the pre-trading survey because it is treated as if the subject "did not know what to enter for this question." We use the answers of the pre-trading survey to replace the "miss-reported" answers in the post-trading survey because it is treated as if the subject "did not know what to update for their beliefs."

- 4.90% (1.20%) of the observations from inexperienced (experienced) sessions are corrected in the above way.

- Simply omitting the above "miss-reported" observations is also performed, and both omitting and correcting the "miss-reported" data do not alter the results.

- Three sessions data are not included in the analysis:

  - One of the sessions was canceled because one player lost internet connection in the middle of the game.

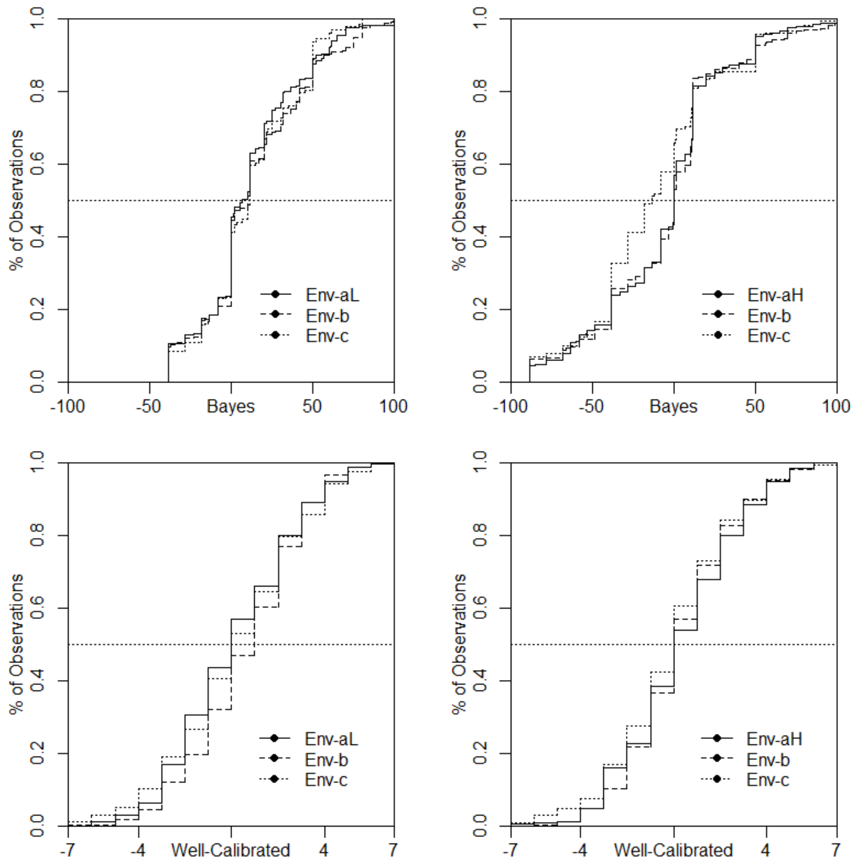
  - In one session, our data analysis indicates one of the players, regardless of the private signal, sold all his/her assets at low prices at the beginning of the period and did not do anything else.

  - In another session, one player's elicited data of question 1 and question 2 are negatively correlated, indicating he/she is willing to sell the asset at a lower price when he/she believes the true state is more likely to be the good state.

  - The two players' earnings are only approximately half of the other players' average earnings. We believe these two players did not understand the game well, and their trading behaviors impacted the markets. The market transaction prices in the two sessions are lower than 200 throughout the session. Therefore, we believe including these three sessions' data does not help answer the research questions.

C.1.2 Overconfidence Across Environments

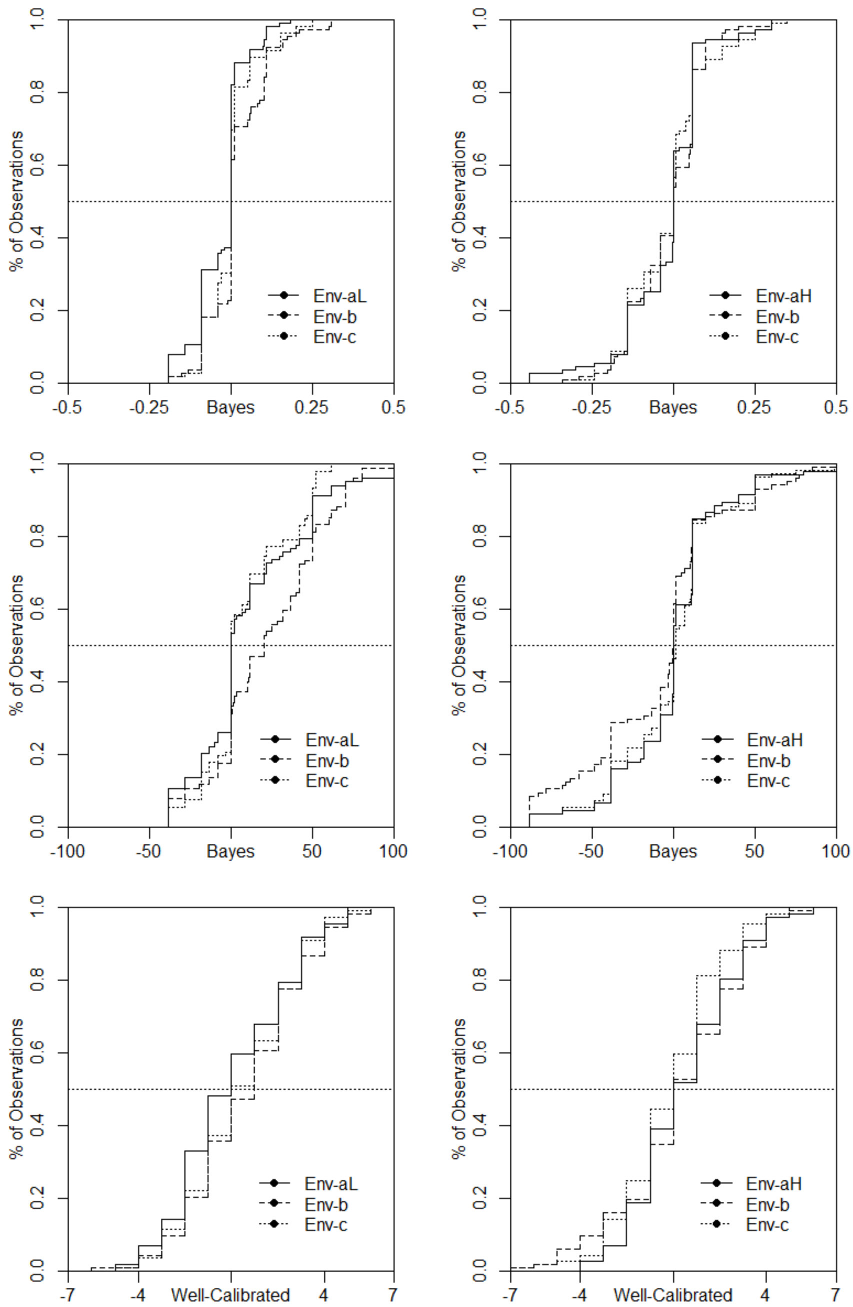
C.1.2.1 Inexperienced Sessions



Note: Figure C1 replicates figure 3.4 using elicited data of question 2 (top) and question 3 (bottom) from inexperienced sessions. Similar to figure 3.4, the left (right) panel is the L-(H-)prcn signal precision. Within each panel, the farther away to the left (right), the more under-(over-)confident the trader is.

Figure C1: Q2 and Q3 Pretrade Over-/Under-confidence

C.1.2.2 Experienced Sessions



Note: Figure C2 replicates figure 3.4 using elicited data of question 1, question 2, and question 3 from experienced sessions.

Figure C2: Pretrade Over-/Under-confidence

C.1.3 Overconfidence Post-Trading (Question 3 Only)

The following two figures plot the CDF of traders' Over-/Under-confidence after trading using only elicited beliefs of question 3. Figure C3 and figure C4 plot graphs for inexperienced and experienced sessions, respectively. The patterns show that there is no significant difference in post-trading confidence level across environments (See hypothesis test results in table C2). The plot on the left (right) panel shows L-prcn (H-prcn).

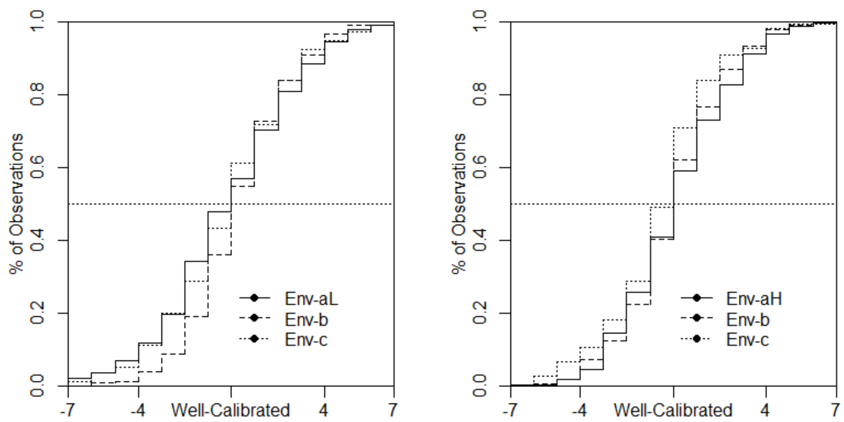


Figure C3: Posttrade Over-/Under-confidence (Inexperienced Sessions)

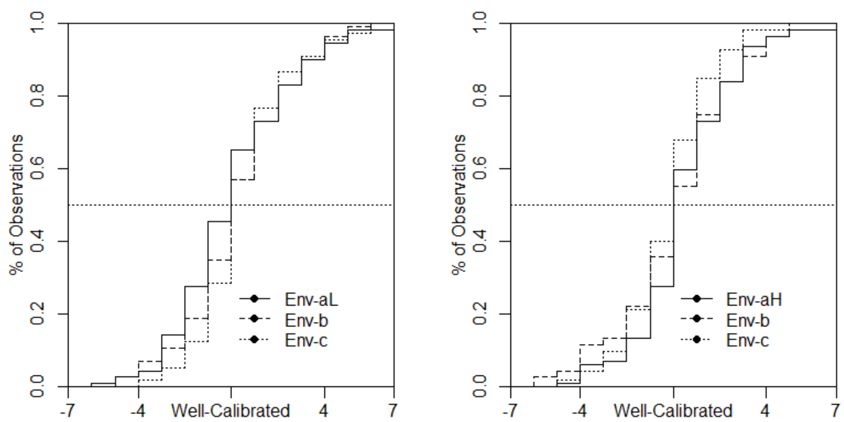


Figure C4: Posttrade Over-/Under-confidence (Experienced Sessions)

C.1.4 Overconfidence Pre and Post Trading (Experienced Sessions)

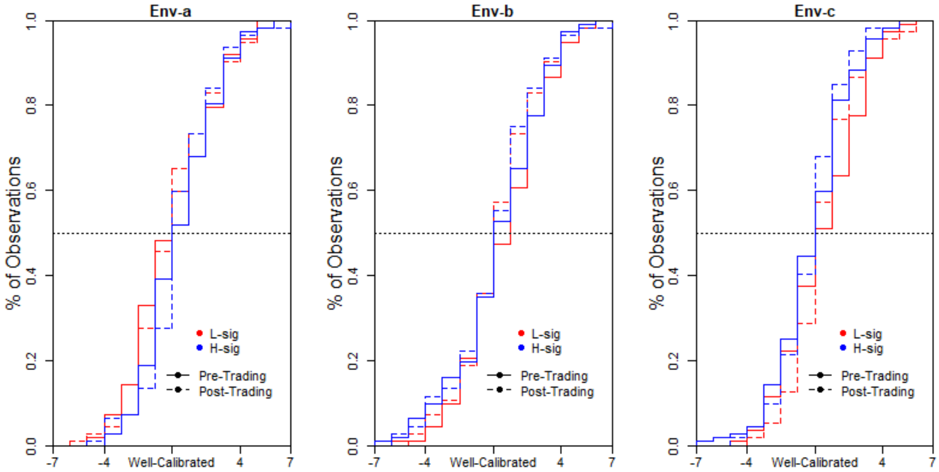


Figure C5: Pre vs Post-Trading Overconfidence (Question 3 in Experienced Sessions)

C.1.5 Level Of Confidence (Inexperienced Sessions)

The level of confidence is defined by the percentage deviations of traders' elicited beliefs from Bayesian P, from V^*, and from "actual rankings". For question 2 and question 3, we rescale the deviation by dividing their x by the possible range of the answers, i.e., $x/200$ for Q2 and $x/8$ for Q3. In this case, the levels of all three questions fall into the same interval of $[-0.5, 0.5]$. We compare the levels conditional on different signal precision. The following two figures illustrate the pattern of inexperienced sessions.

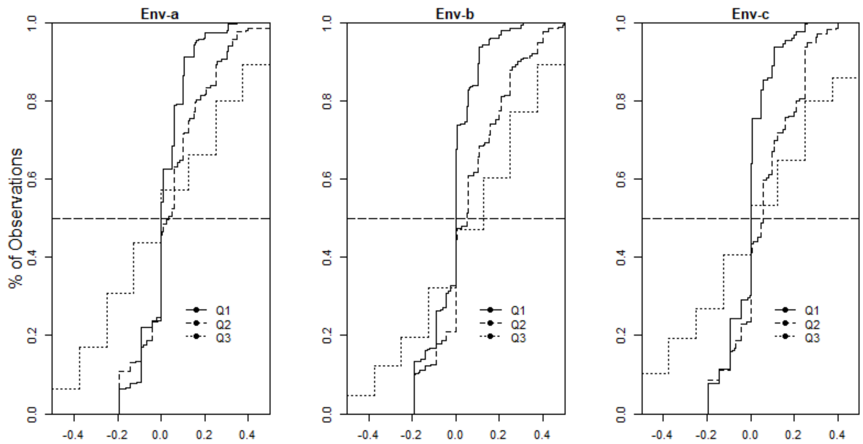


Figure C6: Level of Confidence (L-prcn)

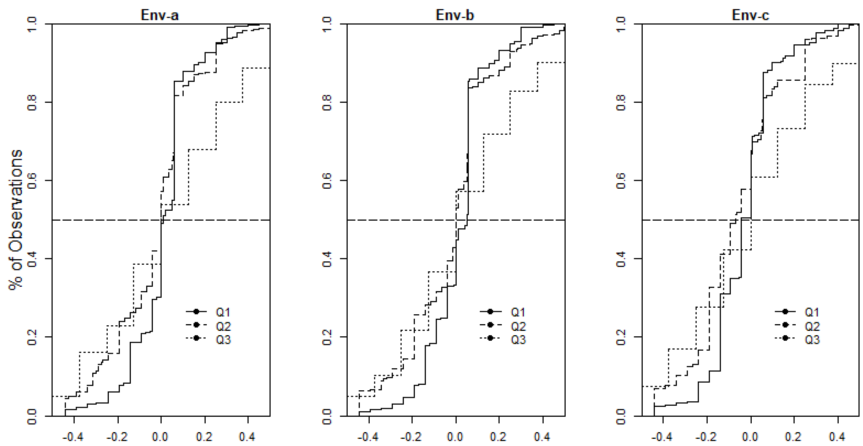


Figure C7: Level of Confidence (H-prcn)

Overall, the level of over and underconfidence are stronger in Q3 than the other two questions. When comparing Q1 and Q2 confidence levels, signal precision seems to have an effect. When traders hold L precision signals, traders are more overconfident in Q2 than Q1 (consistent with Fan et al. (2021)), but exhibit similar level of underconfidence. However, the patterns are reversed when they hold H precision signals. The patterns are similar in experienced sessions, which are shown in Appendix C.1.6.

C.1.6 Level Of Confidence (Experienced Sessions)

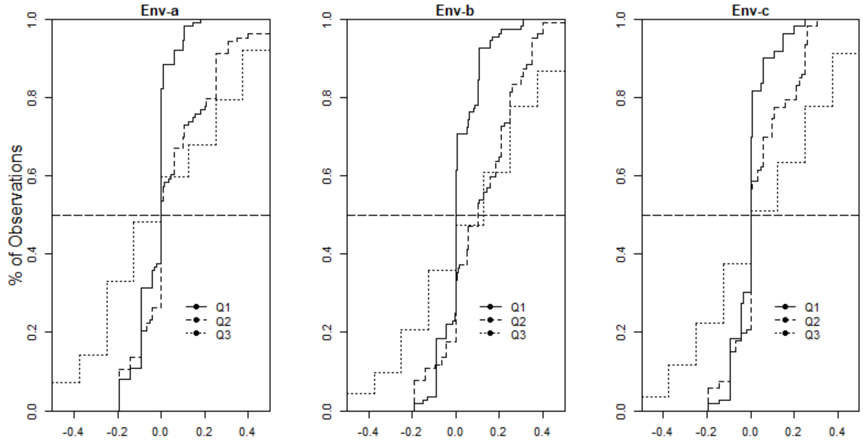


Figure C8: Level of Confidence (L-prcn)

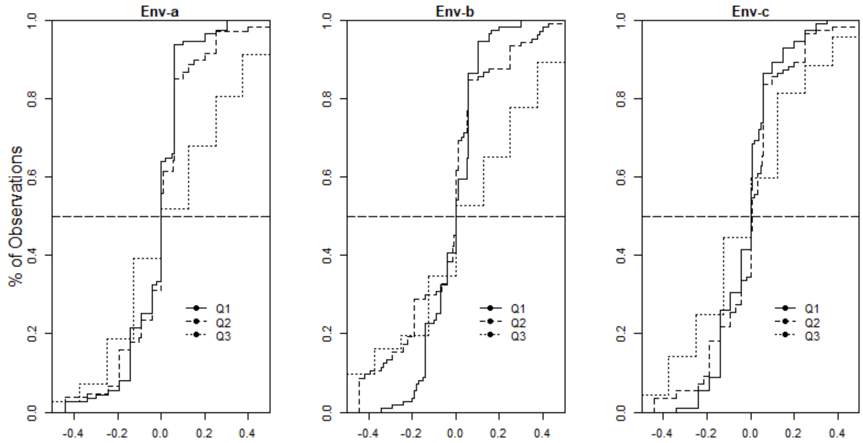
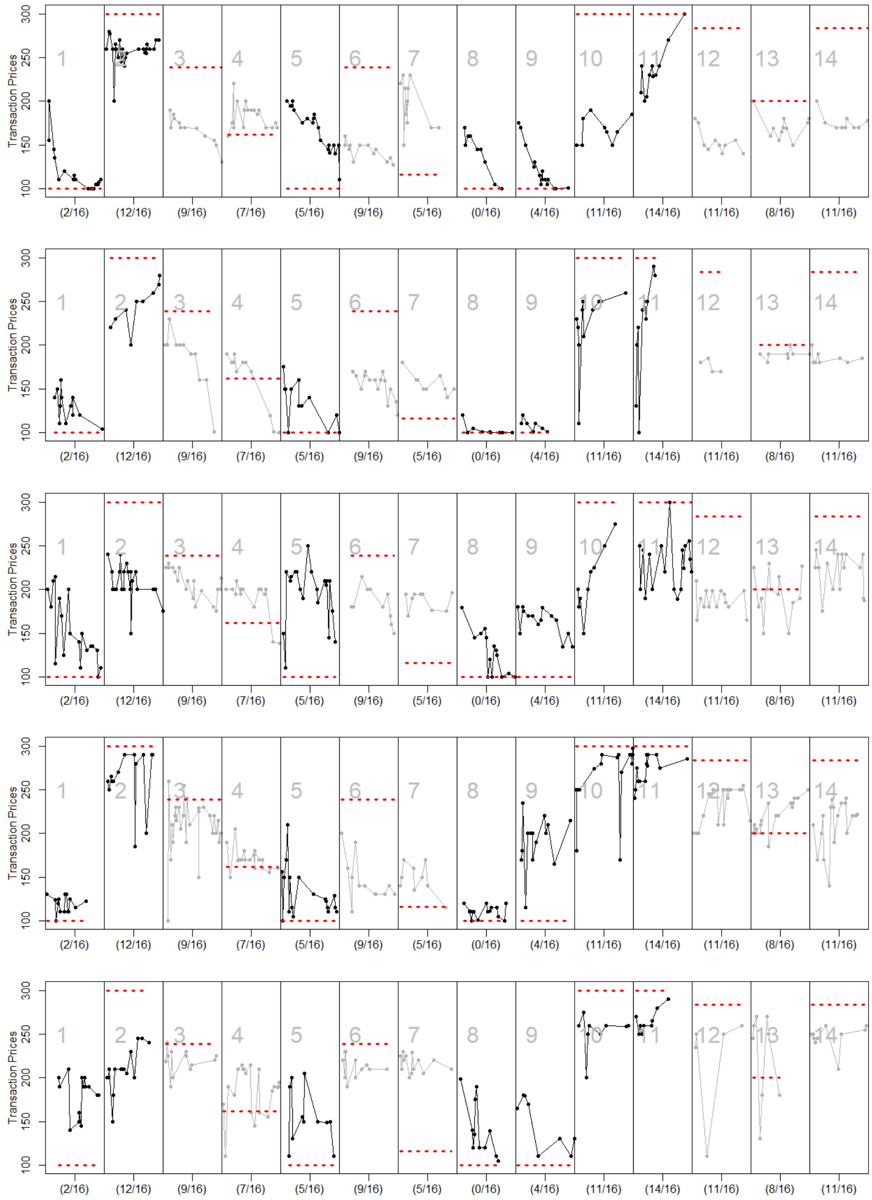


Figure C9: Level of Confidence (H-prcn)

C.1.7 Transaction Prices, Env-a



Note: Transaction Prices in black (gray) means this period is in environment aL (aH).

Figure C10: Transaction Prices Under Env-a (All 5 sessions)

C.1.8 Transaction Prices, Env-b

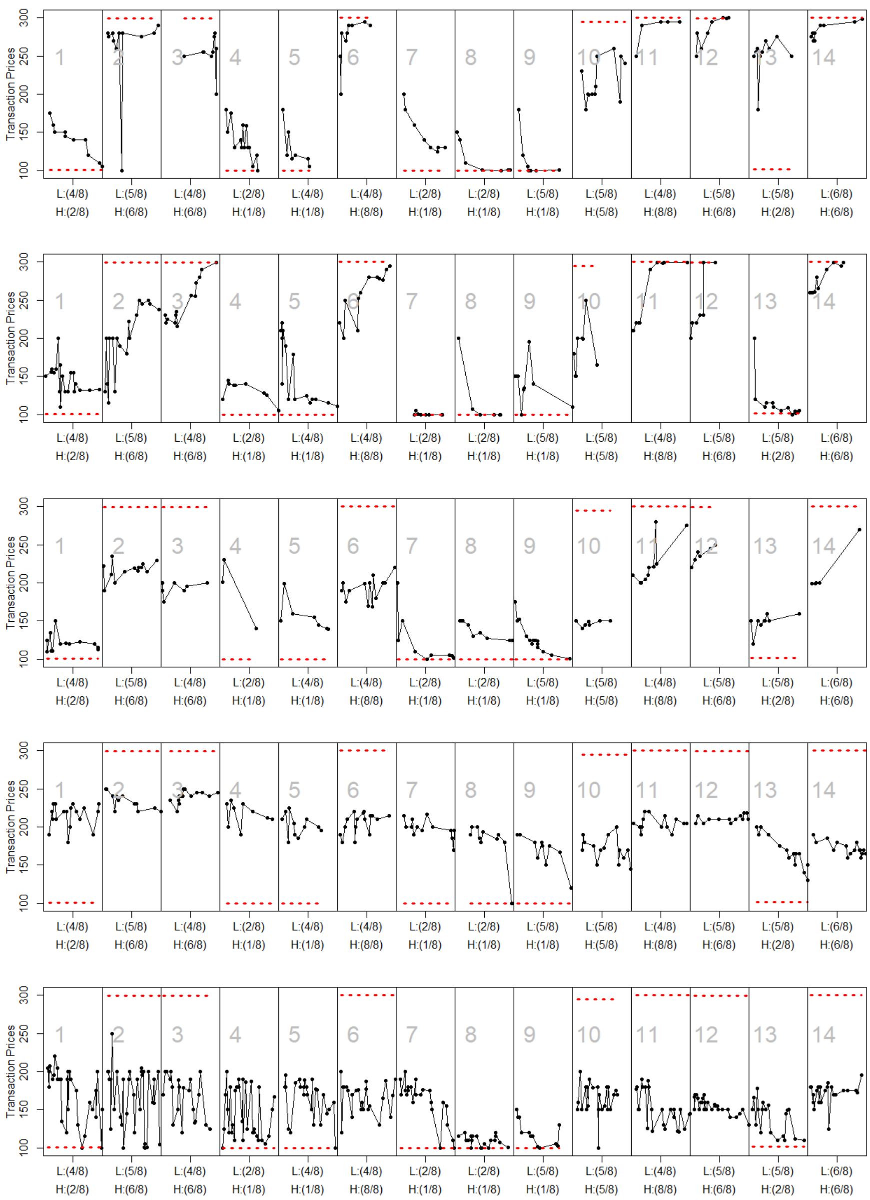


Figure C11: Transaction Prices Under Env-b (All 5 sessions)

C.1.9 Transaction Prices, Env-c

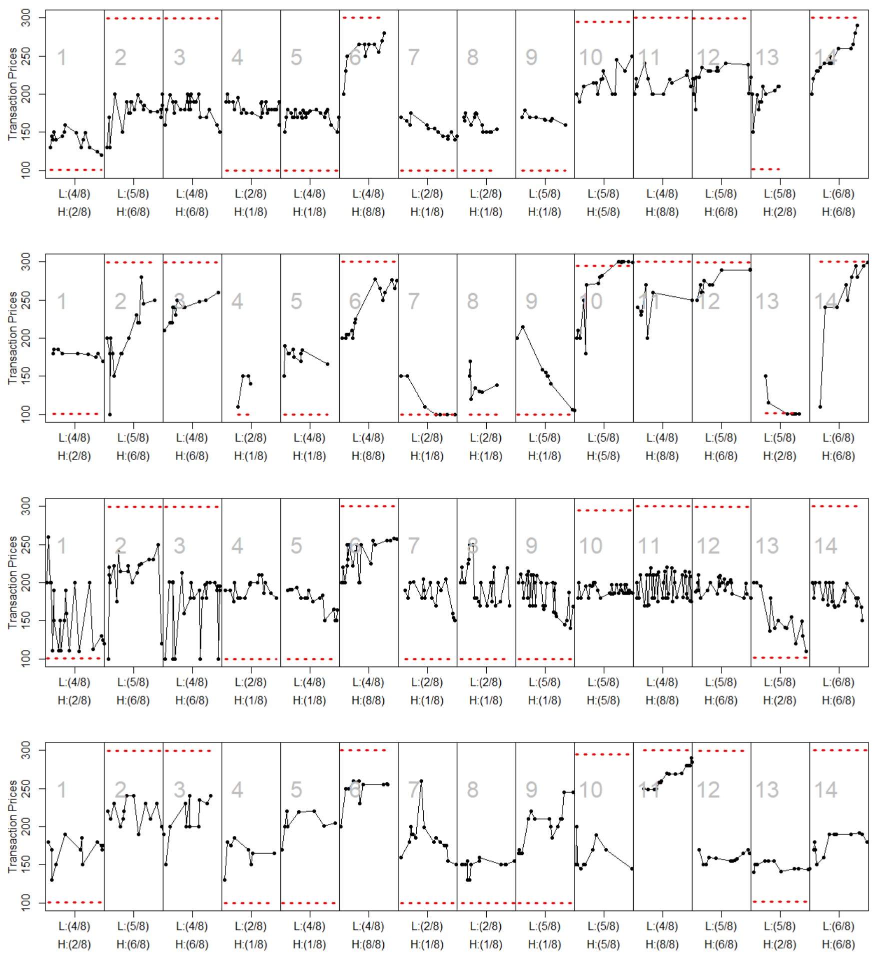


Figure C12: Transaction Prices Under Env-c (All 4 sessions)

C.1.10 Transaction Prices, Experienced Sessions

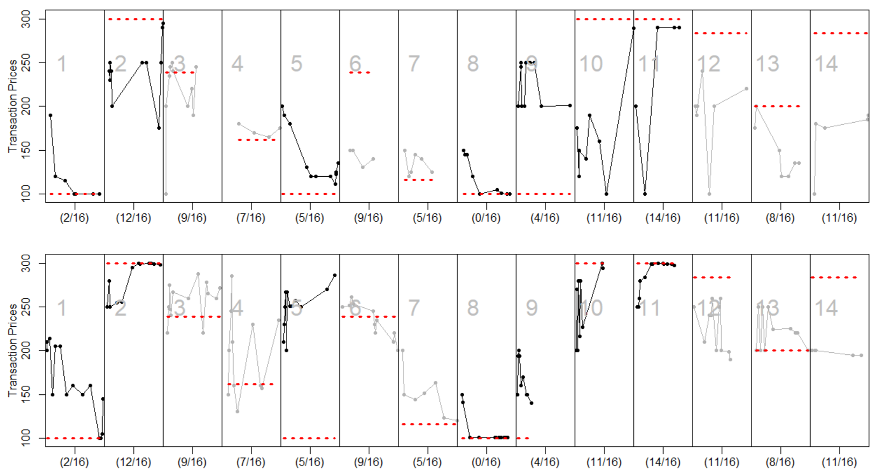


Figure C13: Transaction Prices Under Env-a (Two sessions)

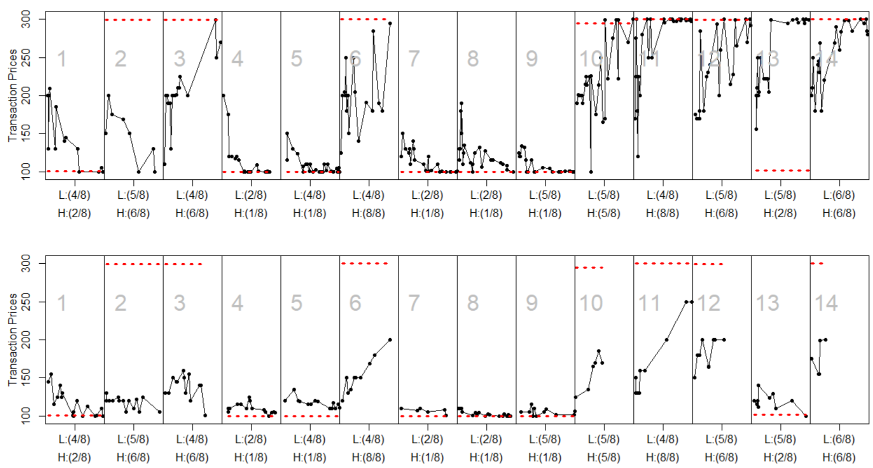


Figure C14: Transaction Prices Under Env-b (Two sessions)

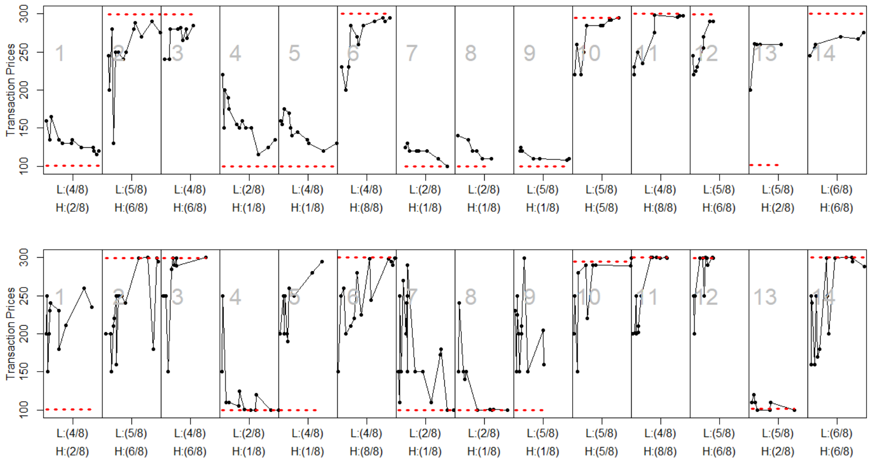


Figure C15: Transaction Prices Under Env-c (Two sessions)

C.1.11 Information Aggregation

C.1.11.1 Estimation in Environment-a

After rescaling the prices into the interval [0,1], the fully aggregated Bayes expected asset value is simply the Bayesian posterior (P) of state G given the realized draws for all 8 subjects:

$$P = Pr(G|S_N) = \frac{q^K \cdot (1-q)^{N-K}}{q^K \cdot (1-q)^{N-K} + q^{N-K} \cdot (1-q)^K},$$

where S_N is the set of all available signals, thus $N = 16$. $q = 0.6(0.8)$ if the period is L-prcn (H-prcn) period. K is the total number of realized black balls in each period.

The model from Page and Siemroth (2021), however, assumes the market can incorporate a subset of the available signals into the prices. The observed prices are set as if the market incorporates a random drawn (without replacement) subset $S_n \subseteq S_N$. Therefore, the observed prices only incorporate $n \leq N$ of all the signals and correspondingly $k \leq K$ number of black balls:

$$\hat{P} = Pr(G|S_n) = \frac{q^k \cdot (1-q)^{n-k}}{q^k \cdot (1-q)^{n-k} + q^{n-k} \cdot (1-q)^k},$$

where, $n = \lceil \lambda N \rceil$ and $\lambda \in [0,1]$ is the proportion of all available signals used and it is the main parameter of interest. $\lceil x \rceil$ is the ceiling function. In order to estimate λ, the model assumes that the observed prices (denoted as P_m for trading periods $m = 1, 2, ..., M$) are generated by:

$$P_m = \hat{P}(\lambda, k) + \epsilon_m \iff \epsilon_m = P_m - \frac{q^k \cdot (1-q)^{\lceil \lambda N \rceil - k}}{q^k \cdot (1-q)^{\lceil \lambda N \rceil - k} + q^{\lceil \lambda N \rceil - k} \cdot (1-q)^k}, \quad \text{(C.1)}$$

where the stochastic deviation ϵ_m is assumed to be normal: $\epsilon_m \sim \mathcal{N}(0, \sigma^2)$, and σ is also a parameter to be estimated. Then the model assigns a probability of observing a specific market price P_m given (λ, k, σ):

$$Pr(P_m|\lambda, k, \sigma) = \phi\Big(\frac{\epsilon_m}{\sigma}\Big)/\sigma = \phi\Big(\frac{P_m - \hat{P}(\lambda, k)}{\sigma}\Big)/\sigma,$$

where $\phi(x)$ is the standard normal density. Also, the probability of drawing a specific k given λ is:

$$Pr(k|\lambda) = \frac{\binom{K_m}{k}\binom{N-K_m}{\lceil \lambda N \rceil - k}}{\binom{N}{\lceil \lambda N \rceil}}$$

Now, given all the above components, the likelihood of observing market price P_m given the model is the probability of drawing an information subset with k black balls and an error term ϵ_m such that $P_m = \hat{P}(\lambda, k) + \epsilon_m$. Thus,

$$\begin{aligned} Pr(P_m|\lambda, \sigma) &= \sum_{k=max\{0, \lceil \lambda N \rceil - N + K_m\}}^{k=min\{\lceil \lambda N \rceil, K_m\}} Pr(P_m|\lambda, k, \sigma) \cdot Pr(k|\lambda) \\ &= \sum_{k=max\{0, \lceil \lambda N \rceil - N + K_m\}}^{k=min\{\lceil \lambda N \rceil, K_m\}} \phi\left(\frac{P_m - \frac{q^k \cdot (1-q)^{\lceil \lambda N \rceil - k}}{q^k \cdot (1-q)^{\lceil \lambda N \rceil - k} + q^{\lceil \lambda N \rceil - k} \cdot (1-q)^k}}{\sigma}\right)/\sigma \cdot \frac{\binom{K_m}{k}\binom{N-K_m}{\lceil \lambda N \rceil - k}}{\binom{N}{\lceil \lambda N \rceil}} \end{aligned}$$

The objective of this estimation is to find $(\hat{\lambda}, \hat{\sigma})$ that maximize the overall log-likelihood of observing the prices $(p_1, p_2, ..., p_M)$ in M trading periods:

$$(\hat{\lambda}, \hat{\sigma}) = \underset{\lambda \in [0,1], \sigma > 0}{\arg\max} \sum_{m=1}^{M} lnPr(P_m|\lambda, \sigma)$$

Due to the ceiling function in $n = \lceil \lambda N \rceil$, λ and σ are not continuous, and therefore, λ and σ both take values in the interval of $[0, 1]$ in 0.01 steps, meaning that $\lambda \in [0, 1] \cap \{0 + 0.01 \cdot l\}_{l=0,1,...,100}$, and similarly for σ.

C.1.11.2 Estimation in Environment-b and Environment-c

The estimation of λ in the other two environments are different because we have two signal precision in the market. Bayesian posterior is the following:

$$P = Pr(G|S_N) = \frac{0.6^{K_L} \cdot 0.4^{N_L - K_L} \cdot 0.8^{K_H} \cdot 0.2^{N_H - K_H}}{0.6^{K_L} \cdot 0.4^{N_L - K_L} \cdot 0.8^{K_H} \cdot 0.2^{N_H - K_H} + 0.6^{N_L - K_L} \cdot 0.4^{K_L} \cdot 0.8^{N_H - K_H} \cdot 0.2^{K_H}}$$

In each market, the total number of private signals are $N = N_L + N_H = 16$. We have $N_L = N_H = 8$ means both L-prcn and H-prcn provide 8 balls and $K_L(K_H)$ refers to the number of black balls from L-prcn (H-prcn). Like the original model, $n = \lceil \lambda N \rceil$ indicates the number of the private signals incorporated into the price, which is a subset of N. But, in environment b and c, we need to figure out how many of n is L-prcn and how many is H-prcn. Since we have half traders receive L-prcn and the other half receive H-prcn, we assume n is splitted evenly between these two precisions. Therefore, we have $n = n_L + n_H$ and $n_L = n_H = \lceil \frac{\lambda}{2} N \rceil$. Therefore, we would have Bayes posterior given the subset of signals equal to:

$$\hat{P} = Pr(G|\lambda, k_L, k_H) = \frac{0.6^{k_L} \cdot 0.4^{n_L - k_L} \cdot 0.8^{k_H} \cdot 0.2^{n_H - k_H}}{0.6^{k_L} \cdot 0.4^{n_L - k_L} \cdot 0.8^{k_H} \cdot 0.2^{n_H - k_H} + 0.6^{n_L - k_L} \cdot 0.4^{k_L} \cdot 0.8^{n_H - k_H} \cdot 0.2^{k_H}}$$

and the probability of observing the market price P_m given the parameters is:

$$Pr(P_m|\lambda, k_L, k_H, \sigma) = \phi\Big(\frac{\epsilon_m}{\sigma}\Big)/\sigma = \phi\Big(\frac{P_m - \hat{P}(\lambda, k_L, k_H)}{\sigma}\Big)/\sigma$$

Similar to the original model, we can find the upper and lower bound for k_L and k_H:

$k_L \in [max\{0, n_L - 8 + K_{L,m}\}, min\{n_L, K_{L,m}\}]$, and $k_H \in [max\{0, n_H - 8 + K_{H,m}\}, min\{n_H, K_{H,m}\}]$

The probability of drawing a specific k_L and k_H given λ is:

$$Pr(k_L|\lambda) = \frac{\binom{K_{L,m}}{k_L}\binom{8-K_{L,m}}{\lceil\frac{\lambda}{2}N\rceil - k_L}}{\binom{8}{\lceil\frac{\lambda}{2}N\rceil}}, \text{ and } Pr(k_H|\lambda) = \frac{\binom{K_{H,m}}{k_H}\binom{8-K_{HH,m}}{\lceil\frac{\lambda}{2}N\rceil - k_H}}{\binom{8}{\lceil\frac{\lambda}{2}N\rceil}}$$

The modified version of the model looks like:

$$\begin{aligned}
Pr(P_m|\lambda, \sigma) &= \sum_{k_H}\sum_{k_L} Pr(P_m|\lambda, k_L, k_H, \sigma) \cdot Pr(k_L|\lambda) \cdot Pr(k_H|\lambda) \\
&= \sum_{k_H}\sum_{k_L} \phi\left(\frac{P_m - \hat{P}_m(\lambda, k_L, k_H)}{\sigma}\right)/\sigma \cdot Pr(k_L|\lambda) \cdot Pr(k_H|\lambda)
\end{aligned}$$

Again, the objective is to find $(\hat{\lambda}, \hat{\sigma})$ that maxmize the overall log-likelihood of observing the prices in M markets,

$$(\hat{\lambda}, \hat{\sigma}) = \underset{\lambda \in [0,1], \sigma > 0}{\arg\max} \sum_{m=1}^{M} lnPr(P_m|\lambda, \sigma)$$

Data	In-experienced λ	Experienced λ	Inexperienced ψ	Experienced ψ
Env_aL	0.06	0.37	0.02	0.25
	[0.00,0.25]	[0.00,0.50]	[-0.19,0.24]	[-0.15,0.54]
Env_aH	0.12	0.25	0.72	0.68
	[0.12,0.31]	[0.12,0.34]	[0.60,0.82]	[0.43,0.89]
Env_b	0.12	0.25	0.48	0.52
	[0.12,0.25]	[0.12,0.37]	[0.39,0.60]	[0.29,0.77]
Env_c	0.12	0.25	0.34	0.73
	[0.00,0.12]	[0.12,0.50]	[0.22,0.44]	[0.55,0.91]

Note: This table replicates Table 3.5 using the very last transaction price in each period.

Table C1: Estimated λ and ψ (Last Price)

C.1.12 Belief Errors Biases

C.1.12.1 Inexperienced Sessions

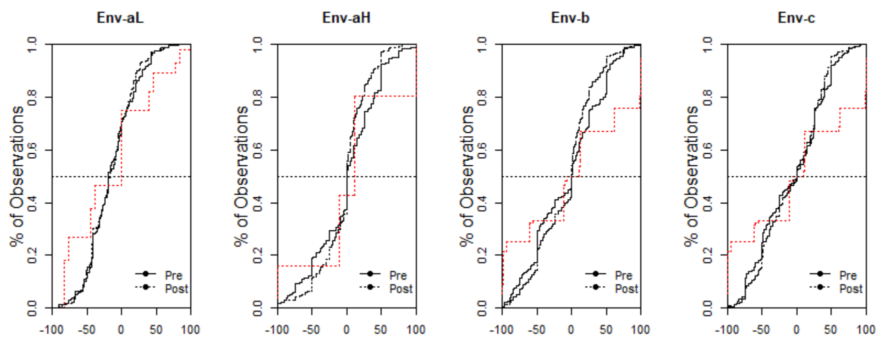


Figure C16: Belief Errors (Question 2)

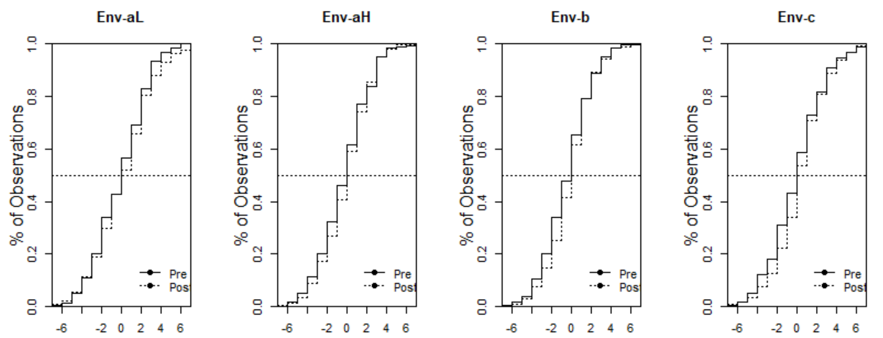


Figure C17: Belief Errors (Question 3)

C.1.12.2 Experienced Sessions

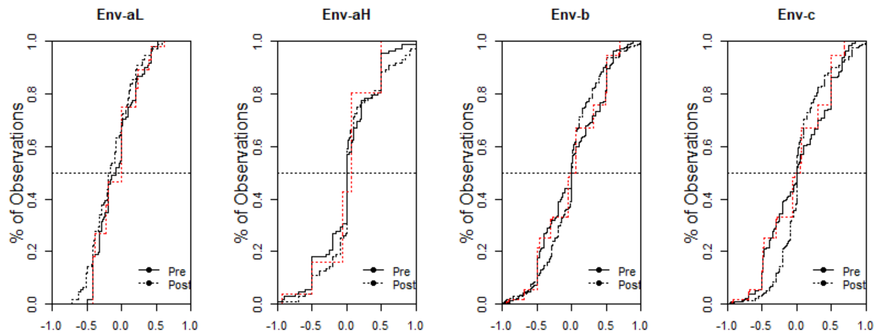


Figure C18: Belief Errors (Question 1)

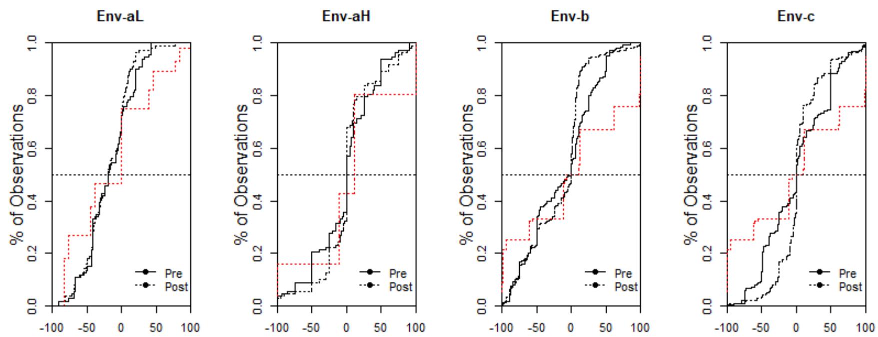


Figure C19: Belief Errors (Question 2)

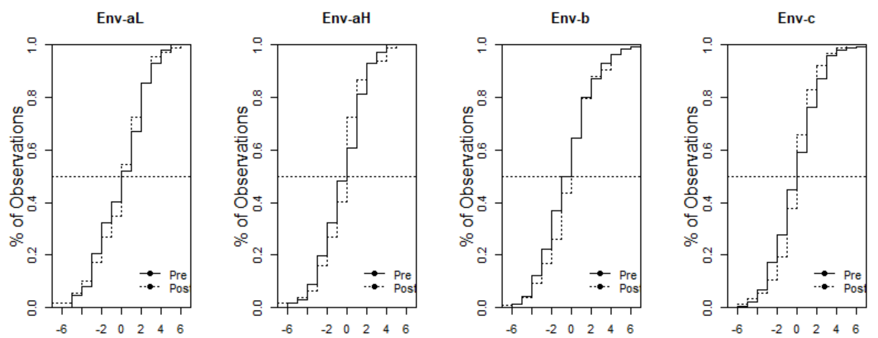


Figure C20: Belief Errors (Question 3)

C.1.13 Wilcoxon Signed-Rank (WSR) Test

	Inexperienced Sessions							
	Question 1		Question 2		Question 3		Question 3 (Posttrade)	
	L-sig	H-sig	L-sig	H-sig	L-sig	H-sig	L-sig	H-sig
H_0 : Env-a = Env-b	0.0024	0.7746	0.3708	0.8372	0.0172	0.8431	0.0444	0.7162
H_0 : Env-b = Env-c	0.4668	0.0000	0.9722	0.0102	0.1019	0.1744	0.0615	0.0149
H_0 : Env-a = Env-c	0.0208	0.0001	0.4447	0.0142	0.6280	0.1456	0.9447	0.0087
	Experienced Sessions							
	Question 1		Question 2		Question 3		Question 3 (Posttrade)	
	L-sig	H-sig	L-sig	H-sig	L-sig	H-sig	L-sig	H-sig
H_0 : Env-a = Env-b	0.0003	0.7284	0.0056	0.0746	0.0603	0.9503	0.2217	0.5423
H_0 : Env-b = Env-c	0.1451	0.5232	0.0014	0.0487	0.6064	0.0813	0.6334	0.1847
H_0 : Env-a = Env-c	0.0165	0.8146	0.9740	0.8001	0.1593	0.0678	0.0593	0.0289

Note: This is the WSR test for figure 3.4 and figures in Appendix C1, Appendix C2, and Appendix C.1.3. Numbers in the table are the p-values.

Table C2: WSR Test on Confidence Levels

	Inexperienced Sessions			
	Env-aL	Env-aH	Env-b	Env-c
Question 1	0.3115	0.7009	0.4794	0.0036
Question 2	0.5846	0.3143	0.4132	0.7749
Question 3	0.3272	0.2747	0.0270	0.0202
	Experienced Sessions			
	Env-aL	Env-aH	Env-b	Env-c
Question 1	0.0909	0.6234	0.9802	0.3022
Question 2	0.6002	0.9655	0.0903	0.4093
Question 3	0.9735	0.9734	0.1975	0.8672

Note: This is the WSR test for figure 3.5 and figures in Appendix C.1.12. Numbers in the table are the p-values.

Table C3: WSR Test on Belief Errors Biases

	In-experienced Sessions					
	Env-a		Env-b		Env-c	
	L-sig	H-sig	L-sig	H-sig	L-sig	H-sig
H_0 : Pre = Post	0.3272	0.2747	0.0714	0.1744	0.1858	0.0489
	Experienced Sessions					
	Env-a		Env-b		Env-c	
	L-sig	H-sig	L-sig	H-sig	L-sig	H-sig
H_0 : Pre = Post	0.9735	0.9734	0.2834	0.4662	0.8972	0.8722

Note: This is the WSR test for figure 3.7 and figures in Appendix C.1.4. The null hypothesis (Pre=Post) is that traders' overconfidence does not change significantly after trading in the market. Numbers in the table are the p-values.

Table C4: WSR Test on Confidence Levels Pre and Post Trading (Question 3 Only)

C.1.14 Overconfidence on Market Prices

The following two tables replicate table 3.6 by using x measured by Q2 and Q3.

	Inexperienced				Experienced			
	Last One		Last Two		Last One		Last Two	
VARIABLES	Q2	se	Q2	se	Q2	se	Q2	se
β_c	0.196	(0.491)	0.0583	(0.442)	-0.521	(0.714)	-0.731	(0.747)
β_{aL}	-0.589	(0.892)	-0.425	(0.810)	1.401	(1.209)	1.569	(1.304)
β_{aH}	-0.857	(0.602)	-0.693	(0.527)	-0.850	(0.889)	-0.802	(0.918)
β_b	0.376	(0.768)	0.607	(0.711)	2.638**	(1.207)	2.870**	(1.148)
Constant	-48.35***	(15.65)	-50.45***	(14.63)	-38.18***	(12.05)	-48.01***	(14.06)
Observations	196		196		84		84	
R-squared	0.247		0.259		0.169		0.177	

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

	Inexperienced				Experienced			
	Last One		Last Two		Last One		Last Two	
VARIABLES	Q3	se	Q3	se	Q3	se	Q3	se
β_c	-10.81	(10.65)	-8.367	(9.812)	10.07	(17.45)	10.58	(17.35)
β_{aL}	20.95	(21.76)	20.91	(20.31)	-60.72	(71.24)	-48.00	(65.72)
β_{aH}	22.79	(16.43)	16.89	(15.16)	17.27	(29.07)	22.70	(28.88)
β_b	18.76	(15.27)	15.97	(14.63)	-16.76	(41.40)	-20.18	(39.92)
Constant	-52.16***	(16.33)	-53.70***	(15.63)	-25.21*	(13.01)	-36.83**	(14.40)
Observations	196		196		84		84	
R-squared	0.236		0.244		0.065		0.062	

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

C.1.15 Overconfidence on Trading Profit

	Inexperienced				Experienced			
VARIABLES	Over	(se)	Under	(se)	Over	(se)	Under	(se)
β_c	-404.2	(275.4)	61.18	(329.4)	-1,064**	(466.0)	-113.3	(646.1)
β_{aL}	340.3	(429.8)	196.5	(569.8)	-462.1	(821.3)	26.13	(915.1)
β_{aH}	-206.8	(332.3)	-332.1	(420.1)	-352.5	(738.6)	-5.203	(731.9)
β_b	-276.0	(337.8)	347.0	(457.4)	3.711	(601.8)	184.9	(899.9)
Constant	477.0***	(64.48)	409.8***	(53.73)	520.3***	(62.35)	423.1***	(93.82)
Observations	498		390		167		173	
R-squared	0.060		0.039		0.102		0.003	

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: This table replicates table 3.7 with x measured by survey question 1.

Table C5: Overconfidence on Trading Profit (x measured by Question 1)

C.1.16 Overconfidence on Taker Volumes

	In-experienced			Experienced		
VARIABLES	Q1	Q2	Q3	Q1	Q2	Q3
β_c	-0.655	0.00700*	0.129**	-0.0494	0.00685**	0.108**
	(0.896)	(0.00394)	(0.0638)	(0.956)	(0.00328)	(0.0530)
β_{aL}	0.310	-0.0112**	0.00140	2.806*	-0.00792*	0.0145
	(1.365)	(0.00529)	(0.0766)	(1.517)	(0.00459)	(0.0751)
β_{aH}	3.067***	-0.00255	-0.0248	1.171	-0.00411	0.152
	(1.130)	(0.00472)	(0.0791)	(1.527)	(0.00578)	(0.0967)
β_b	1.657	-0.00511	-0.159**	-1.682	-0.0135**	-0.112
	(1.232)	(0.00455)	(0.0732)	(2.008)	(0.00564)	(0.0934)
Constant	1.727***	1.779***	1.641***	0.972***	1.007***	0.900***
	(0.216)	(0.232)	(0.202)	(0.122)	(0.134)	(0.119)
Observations	1,526	1,464	1,568	663	631	672
R-squared	0.063	0.056	0.073	0.082	0.080	0.089

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table C6: Overconfidence on Taker Volume

C.1.17 Overconfidence on Sell-Buy-Gap

	Inexperienced				Experienced			
VARIABLES	Maker	se	Taker	se	Maker	se	Taker	se
β_c	-0.0531	(0.0776)	-0.220**	(0.109)	-0.0115	(0.216)	-0.250	(0.545)
β_{aL}	0.296	(0.206)	-0.0252	(0.198)	0.0533	(0.279)	0.222	(0.719)
β_{aH}	-0.140	(0.173)	0.108	(0.188)	-0.482	(0.318)	0.234	(0.617)
β_b	0.0160	(0.104)	0.162	(0.130)	-0.121	(0.247)	0.146	(0.566)
Constant	15.98***	(5.299)	-19.99***	(6.289)	38.10**	(14.58)	-86.52**	(36.56)
Observations	362		266		144		105	
R-squared	0.084		0.097		0.254		0.127	

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table C7: Overconfidence on Sell-Buy-Gap (Question 2)

C.1.18 Summary of Buys and Sells

C.1.18.1 Inexperienced Sessions

	# of Sell: 0	# of Sell: 1	# of Sell: 2	# of Sell: More	Sum of Obs.
# of Buy: 0	13	20	28	0	61
# of Buy: 1	27	32	34	21	114
# of Buy: 2	15	14	8	13	50
# of Buy: More	15	5	10	25	55
Sum of Obs.	70	71	80	59	

Table C8: Summary of Buys and Sells (Env-aL)

	# of Sell: 0	# of Sell: 1	# of Sell: 2	# of Sell: More	Sum of Obs.
# of Buy: 0	11	19	48	0	78
# of Buy: 1	14	19	25	32	90
# of Buy: 2	12	10	15	12	49
# of Buy: More	23	11	10	19	63
Sum of Obs.	60	59	98	63	

Table C9: Summary of Buys and Sells (Env-aH)

	# of Sell: 0	# of Sell: 1	# of Sell: 2	# of Sell: More	Sum of Obs.
# of Buy: 0	31	35	124	0	190
# of Buy: 1	36	31	40	30	137
# of Buy: 2	34	18	12	38	102
# of Buy: More	52	13	13	53	131
Sum of Obs.	153	97	189	121	

Table C10: Summary of Buys and Sells (Env-b)

	# of Sell: 0	# of Sell: 1	# of Sell: 2	# of Sell: More	Sum of Obs.
# of Buy: 0	24	38	100	0	162
# of Buy: 1	22	23	29	27	101
# of Buy: 2	16	15	7	17	55
# of Buy: More	39	14	15	62	130
Sum of Obs.	101	90	151	106	

Table C11: Summary of Buys and Sells (Env-c)

C.1.18.2 Experienced Sessions

	# of Sell: 0	# of Sell: 1	# of Sell: 2	# of Sell: More	Sum of Obs.
# of Buy: 0	62	68	126	0	256
# of Buy: 1	63	48	36	38	185
# of Buy: 2	25	16	20	25	86
# of Buy: More	51	21	20	53	145
Sum of Obs.	201	153	202	116	

Table C12: Summary of Buys and Sells (All Envs)

	# of Sell: 0	# of Sell: 1	# of Sell: 2	# of Sell: More	Sum of Obs
# of Buy: 0	18	16	13	0	47
# of Buy: 1	14	13	9	4	40
# of Buy: 2	8	2	1	4	15
# of Buy: More	5	3	0	2	10
Sum of Obs.	45	34	23	10	

Table C13: Summary of Buys and Sells (Env-aL)

	# of Sell: 0	# of Sell: 1	# of Sell: 2	# of Sell: More	Sum of Obs.
# of Buy: 0	14	7	23	0	44
# of Buy: 1	14	5	8	9	36
# of Buy: 2	6	2	2	3	13
# of Buy: More	5	6	5	3	19
Sum of Obs.	39	20	38	15	

Table C14: Summary of Buys and Sells (Env-aH)

	# of Sell: 0	# of Sell: 1	# of Sell: 2	# of Sell: More	Sum of Obs.
# of Buy: 0	15	22	40	0	77
# of Buy: 1	13	15	7	14	49
# of Buy: 2	7	4	6	12	29
# of Buy: More	13	8	10	38	69
Sum of Obs.	48	49	63	64	

Table C15: Summary of Buys and Sells (Env-b)

	# of Sell: 0	# of Sell: 1	# of Sell: 2	# of Sell: More	Sum of Obs.
# of Buy: 0	15	23	50	0	88
# of Buy: 1	22	15	12	11	60
# of Buy: 2	4	8	11	6	29
# of Buy: More	28	4	5	10	47
Sum of Obs.	69	50	78	27	

Table C16: Summary of Buys and Sells (Env-c)

C.1.19 Results Exclude Observations with Bayes Equal to 0.5.

Recall that our current definition of overconfidence says x is always categorized as overconfident and take a positive value when $P = 0.5$. One of the obvious shortcomings of this definition is that this does not allow underconfidence. This subsection presents relevant graphs and tables of x, excluding observations that have $P = 0.5$. We denote these the results as "EXC."

Figure C21 and C22 are measured x across environments excluding $P = 0.5$ cases. The systematic patterns are the following: 1. Overall in inexperienced sessions, elicited beliefs in environment c exhibit more underconfident bias when traders hold H-prcn signals. 2. In experienced sessions, however, traders who hold L-prcn in homogeneous environment (environment a) are relatively more underconfident.

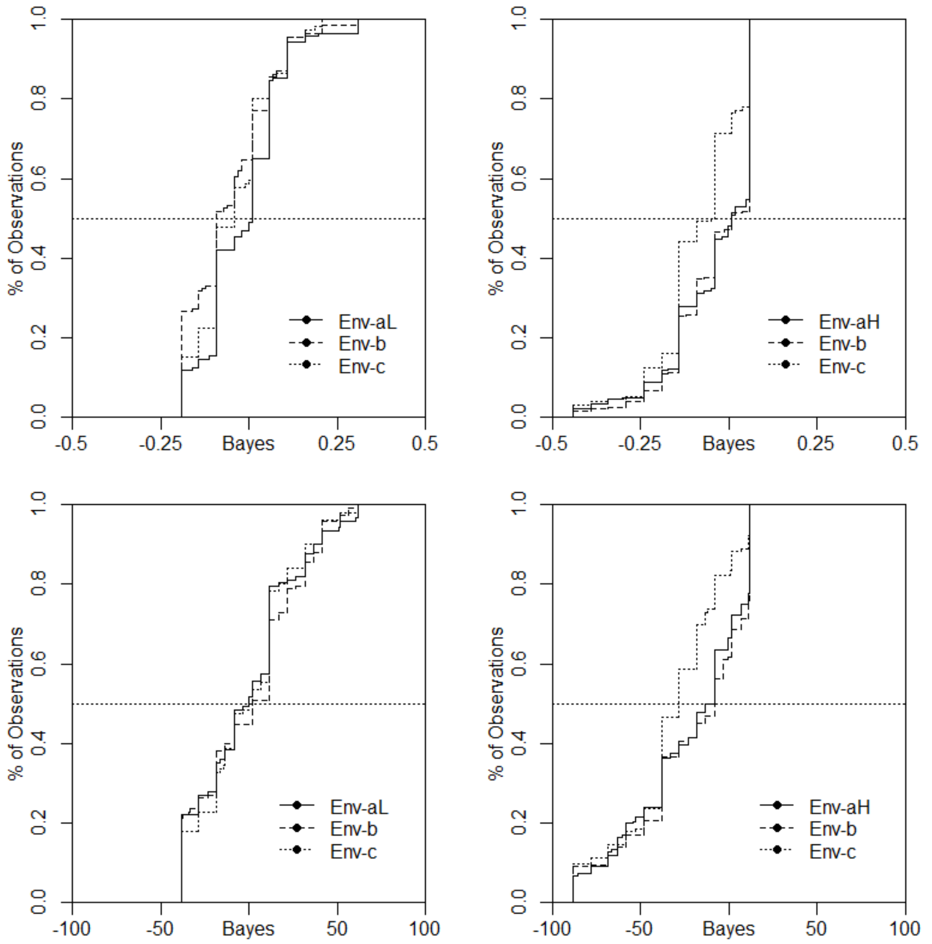


Figure C21: Q1 and Q2 Pretrade Over-/Under-confidence Inexperienced Sessions (EXC)

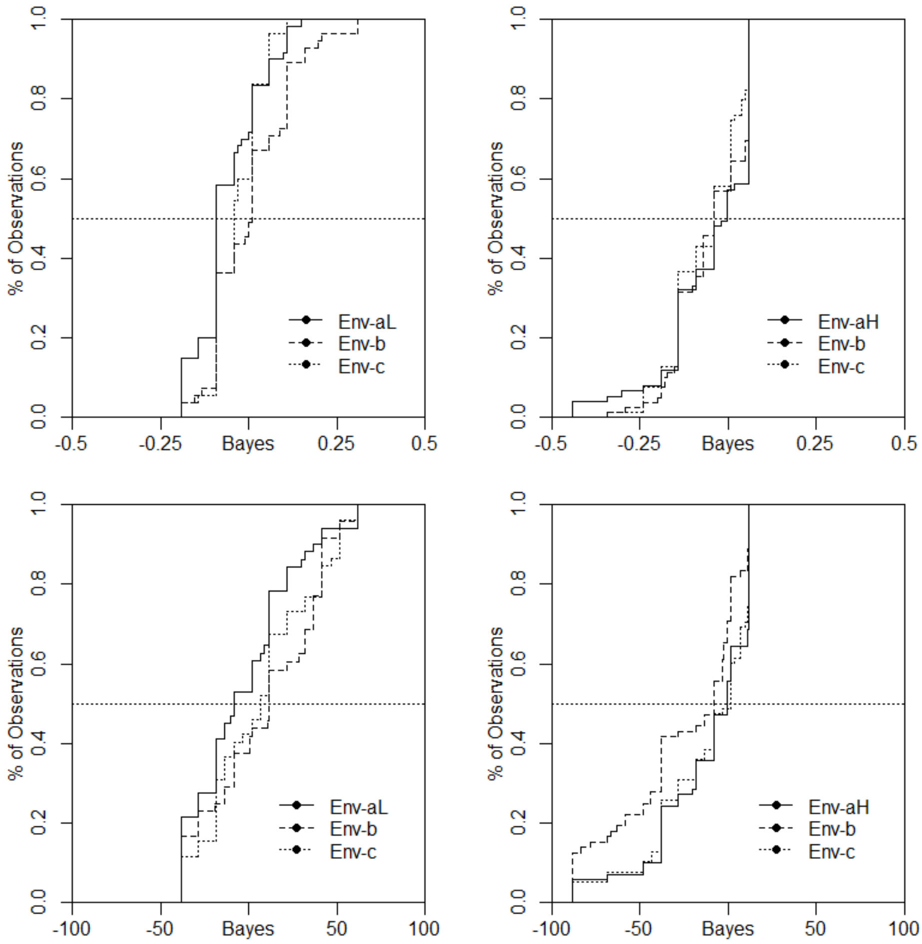


Figure C22: Q1 and Q2 Pretrade Over-/Under-confidence Experienced Sessions (EXC)

Table C17 has a similar pattern as Table 3.2, suggesting correlations of over-confidence are similar with and without the $P = 0.5$ cases.

Inexperienced Sessions					
	All Envs	Env-aL	Env-aH	Env-b	Env-c
Corr(Q1,Q2)	0.41***	0.44***	0.34***	0.47***	0.33***
Corr(Q2,Q3)	0.03	0.02	-0.09	0.04	0.02
Corr(Q1,Q3)	0.01	0.06	-0.07	0.01	0.03
(Obs.)	(820)	(117)	(166)	(294)	(243)
Experienced Sessions					
	All Envs	Env-aL	Env-aH	Env-b	Env-c
Corr(Q1,Q2)	0.21***	0.40***	0.30**	0.13	0.28***
Corr(Q2,Q3)	0.10*	-0.21	-0.00	0.09	0.30***
Corr(Q1,Q3)	0.02	0.17	0.04	-0.04	-0.04
(Obs.)	(365)	(51)	(69)	(117)	(128)

Table C17: Overconfidence Correlations (EXC)

C.1.19.1 Overconfidence on Trading Profit

Table C18 replicates equation 3.8 but exclude $P = 0.5$ cases. Note that, before, we categorize x only as overconfident when $P = 0.5$. Therefore, after dropping them, we only see changes in "Overconfident" columns. So the "underconfident" columns remain the same. The following table suggests that excluding $P = 0.5$ cases does not have a significant impact on result 3.4.4.1.

	Inexperienced				Experienced			
VARIABLES	Over	se	under	se	Over	se	under	se
β_c	-3.547*	(1.823)	2.779***	(0.857)	-7.884***	(2.540)	0.772	(1.295)
β_{aL}	-0.886	(2.818)	-0.329	(1.954)	7.409**	(3.296)	-1.096	(2.661)
β_{aH}	0.0755	(6.049)	-3.870***	(1.146)	0.786	(6.670)	-0.202	(2.072)
β_b	0.226	(2.280)	-0.156	(1.196)	7.004**	(3.005)	3.712**	(1.786)
Constant	594.3***	(102.9)	401.8***	(59.40)	423.4***	(60.52)	443.7***	(104.6)
Observations	334		523		176		195	
R-squared	0.063		0.065		0.170		0.076	

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table C18: Coefficient Estimates for Equation (3.8) (EXC)

Below, table C19 reports the same regression with x measured by Q1. The coefficients are mainly insignificant.

	Inexperienced				Experienced			
VARIABLES	Over	se	Under	se	Over	se	Under	se
β_c	625.4	(661.3)	215.5	(265.1)	1,973	(1,242)	-141.3	(381.4)
β_{aL}	-314.5	(829.5)	328.1	(472.6)	-2,419	(1,508)	-8.492	(594.4)
β_{aH}	106.2	(1,160)	-379.2	(343.6)	-1,361	(1,768)	-62.30	(471.8)
β_b	-1,683**	(815.7)	-156.9	(392.9)	-2,271*	(1,280)	45.38	(613.4)
Constant	421.3***	(88.27)	382.3***	(48.63)	375.4***	(67.85)	452.2***	(62.93)
Observations	406		513		166		237	
R-squared	0.037		0.030		0.032		0.004	

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table C19: Coefficient Estimates for Equation (3.8) (x measured by Q1, EXC)

C.1.19.2 Overconfidence on Sell-Buy-Gap

The following table replicates regression 3.9. It turns out that, after dropping $P = 0.5$ observations, the result is slightly different from result 3.4.4.2, and it supports the following: Takers tend to trade at less favorable prices than do makers. Being more overconfident may improve (resp. worsen) traders' transaction prices in ambiguous environment c (resp. aL and b), but the effect may be partially reversed in environment aH.

	Inexperienced				Experienced			
VARIABLES	Maker	se	Taker	se	Maker	se	Taker	se
β_c	92.13**	(43.53)	31.13	(58.05)	15.44	(48.65)	34.55	(106.7)
β_{aL}	-186.8***	(53.20)	-106.3	(77.05)	-113.2	(90.17)	-106.2	(227.0)
β_{aH}	-103.8	(70.02)	41.88	(84.16)	-103.1	(80.46)	4.139	(153.5)
β_b	-50.42	(52.06)	-48.88	(63.93)	-24.25	(68.17)	-108.3	(115.8)
Constant	17.62**	(7.433)	-15.15*	(7.868)	46.16**	(18.10)	-145.8***	(37.67)
Observations	236		161		92		75	
R-squared	0.113		0.119		0.304		0.184	

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table C20: Overconfidence on Sell-Buy-Gap (EXC)

Table C21 shows regression with x measured by Q2. The patterns are in general not significant and not systematic.

	Inexperienced				Experienced			
VARIABLES	Maker	se	Taker	se	Maker	se	Taker	se
β_c	-0.0597	(0.181)	-0.232	(0.233)	-0.141	(0.231)	0.234	(0.633)
β_{aL}	0.0307	(0.352)	-0.418	(0.375)	0.312	(0.704)	-0.148	(1.054)
β_{aH}	-0.170	(0.303)	0.496*	(0.293)	-0.530	(0.459)	-0.508	(0.736)
β_b	0.0268	(0.200)	0.0616	(0.285)	-0.00774	(0.309)	-0.382	(0.664)
Constant	12.39	(10.23)	-11.95	(9.240)	35.52	(26.07)	-164.0***	(6.858)
Observations	224		138		87		73	
R-squared	0.109		0.208		0.324		0.218	

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table C21: Overconfidence on Sell-Buy-Gap (Question 2) (EXC)

C.1.19.3 Overconfidence On Trading Volume

The following table replicates equation 3.10, but excluding $P = 0.5$. The result is largely the same as result 3.4.4.3.

	Inexperienced			Experienced		
VARIABLES	Q1	Q2	Q3	Q1	Q2	Q3
β_c	0.850	0.00503	-0.00599	-4.028***	-0.00575	-0.0997**
	(0.989)	(0.00315)	(0.0297)	(1.505)	(0.00406)	(0.0438)
β_{aL}	0.215	-0.0140**	-0.0588	5.835**	0.00236	0.0644
	(1.501)	(0.00545)	(0.0491)	(2.394)	(0.00621)	(0.0668)
β_{aH}	-0.0199	0.00125	-0.0981**	5.449***	0.0103	0.0805
	(1.379)	(0.00419)	(0.0484)	(2.087)	(0.00846)	(0.0745)
β_b	-0.103	-0.00199	-0.0250	1.727	0.0134**	0.0954
	(1.388)	(0.00531)	(0.0442)	(2.072)	(0.00611)	(0.0646)
Constant	1.793***	1.946***	1.708***	0.992***	0.958***	0.942***
	(0.207)	(0.210)	(0.132)	(0.142)	(0.137)	(0.0944)
Observations	919	857	1,568	403	371	672
R-squared	0.095	0.117	0.099	0.117	0.119	0.099

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table C22: Overconfidence on Maker Volume (EXC)

Table C23 reports parallel results on taker volume. In general, the patterns are not systematic and less significant.

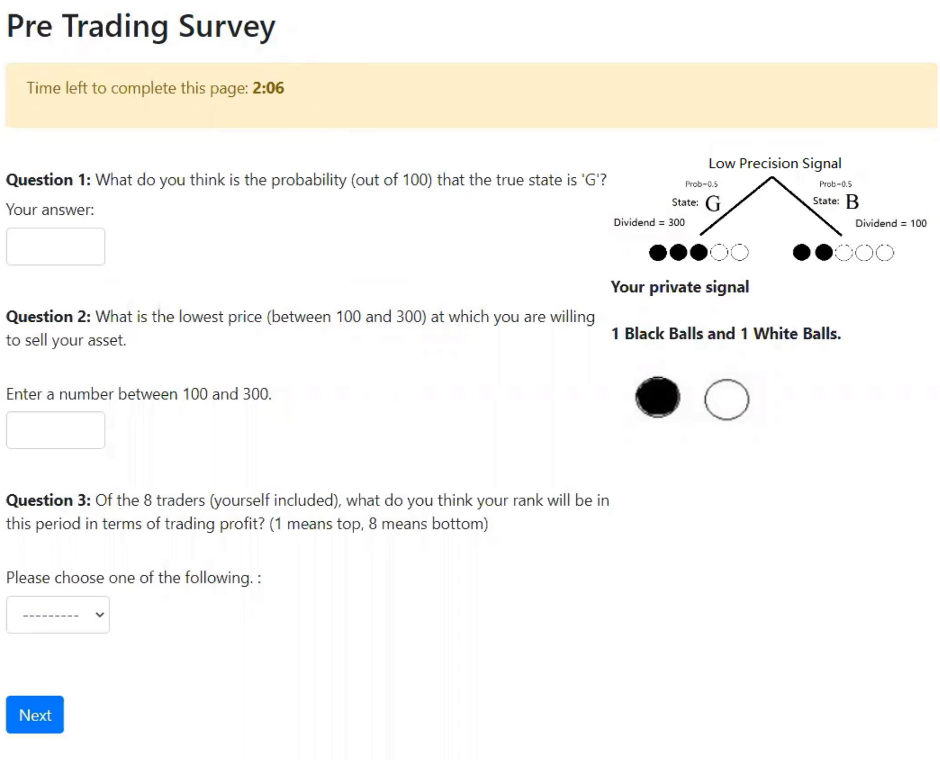
	In-experienced			Experienced		
VARIABLES	Q1	Q2	Q3	Q1	Q2	Q3
β_c	-0.344	-0.000615	0.129**	1.483	0.0137***	0.108**
	(1.496)	(0.00576)	(0.0638)	(1.466)	(0.00435)	(0.0530)
β_{aL}	-0.141	0.00277	0.00140	1.684	-0.0182***	0.0145
	(1.963)	(0.00899)	(0.0766)	(1.911)	(0.00650)	(0.0751)
β_{aH}	3.945**	0.00618	-0.0248	-1.755	-0.0201***	0.152
	(1.841)	(0.00671)	(0.0791)	(1.857)	(0.00663)	(0.0967)
β_b	2.325	0.00319	-0.159**	-4.119*	-0.0218***	-0.112
	(1.848)	(0.00667)	(0.0732)	(2.367)	(0.00741)	(0.0934)
Constant	1.968***	1.973***	1.641***	0.837***	0.807***	0.900***
	(0.340)	(0.358)	(0.202)	(0.124)	(0.151)	(0.119)
Observations	919	857	1,568	403	371	672
R-squared	0.066	0.044	0.073	0.133	0.130	0.089

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table C23: Overconfidence on Taker Volume (EXC)

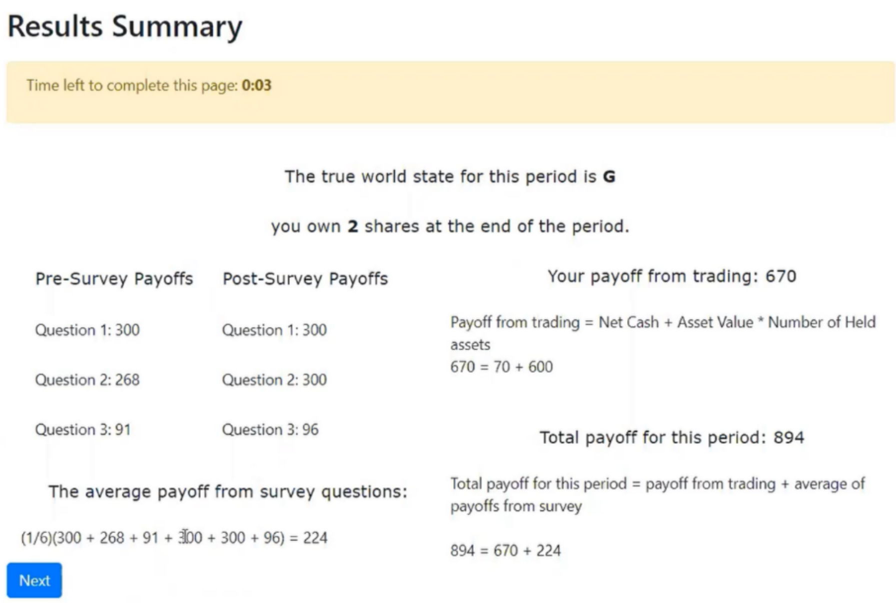
C.2 Experiment Instructions and User Interface

C.2.1 User Interface



Note: In both environment-a and environment-b, we label each trader's signal precision to remind them if their signal precision is a "Low Precision Signal" or a "High Precision Signal" as in the above figure. However, in environment-c, to avoid traders get a sense of how precise their private signals are relative to other traders, we do not label their signal precision, all else stays the same on this UI.

Figure C23: UI of Pre-Trading Survey



Note: This is the UI of the "Results Summary" page, which is the last feedback page that summarizes all the previous feedback pages.

Figure C24: UI of Performance Summary

The other feedback pages are the following. "Your Results" page shows the true state of this period and traders' final assets holdings. "Trading Result" shows traders their payoff from trading, "Survey Result" shows their payoffs from answering each question and the average of the payoffs. "Total Payoff" shows them the total experimental points they earned in this period. These individual feedback pages are all summarized in figure C24 and therefore we do not repeat them in the appendix.

C.2.2 Experiment Instructions

INSTRUCTIONS

Welcome! This is an experiment in the economics of decision-making. From now until the end of the experiment, please turn off your cell phone and do not communicate with other participants. If you pay close attention to the instructions and make good decisions, you can earn a significant sum of money which will be paid in cash at the end of the session. If you have any questions, please do not hesitate to let the experimenter know. We expect and appreciate your cooperation.

The Basic Idea.

You will trade an asset with other participants, using computerized trading screens. The value of the asset depends on the state of the world. If the state is good (G) then the asset pays 300 points to you, but it pays only 100 points if the state is bad (B). Before each trading period, the computer tosses a fair coin, and assigns state G if the coin comes up heads or B if tails. That is, at the beginning of each period there is a 50% likelihood for each state. **All traders trade under the same state of the world**. You will not know what the true state is but, as explained below, you will get a useful hint from the computer in the form of a private signal. After receiving the private signal, you will be asked **how likely** you now think that the true state is G. Then you will be able to **trade with other participants**. Since they also get private signals that may differ from yours, the trading prices you see may also provide hints about the true state and the asset value. After the trading period is over you will be asked about your **final beliefs** about the true state in this period, based on both your private signal and on other traders' offers and trades. Then there will be feedback after the trading revealing the true state of the world, showing your payoff from trading, payoff for providing your beliefs, and the total payoff for this period.

Each trading period lasts approximately 3 minutes. There will be between 12 and 18 trading periods in today's experiment. The total of all points you earn in all periods will be converted into US dollars and paid to you in cash.

Private Signal

There are two types of private signals: Low precision signal (L-sig) and High precision signal (H-sig).

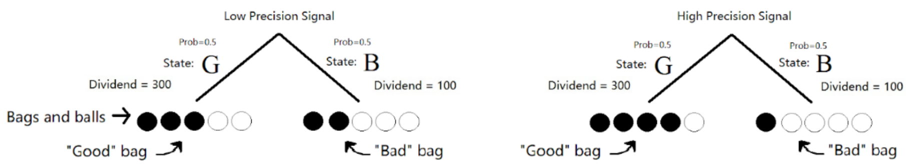


Figure 1

The left side of Figure 1 illustrates L-sig in terms of two bags that each contain 5 balls. If the true state is G then the computer draws balls **only** from a "Good bag" that contains 3 black balls and 2 white balls. Similarly, if the true state is "B", the computer will draw balls **only** from the "Bad bag" that contains only 2 black balls plus 3 white balls. The computer will randomly draw two balls with replacement from the bag corresponding to the true state and tell you the colors of those two balls. That is your private signal.

If your private signal is two black balls, then you should conclude that G is more likely than 50% because those two balls are more likely drawn from the "good bag", which contains more black balls. Likewise, if your private signal is 2 white balls then the true state is more likely to be "B". If your private signal is 1 black ball and 1 white ball, then the probability the true state is G might be a number in between the previous two cases.

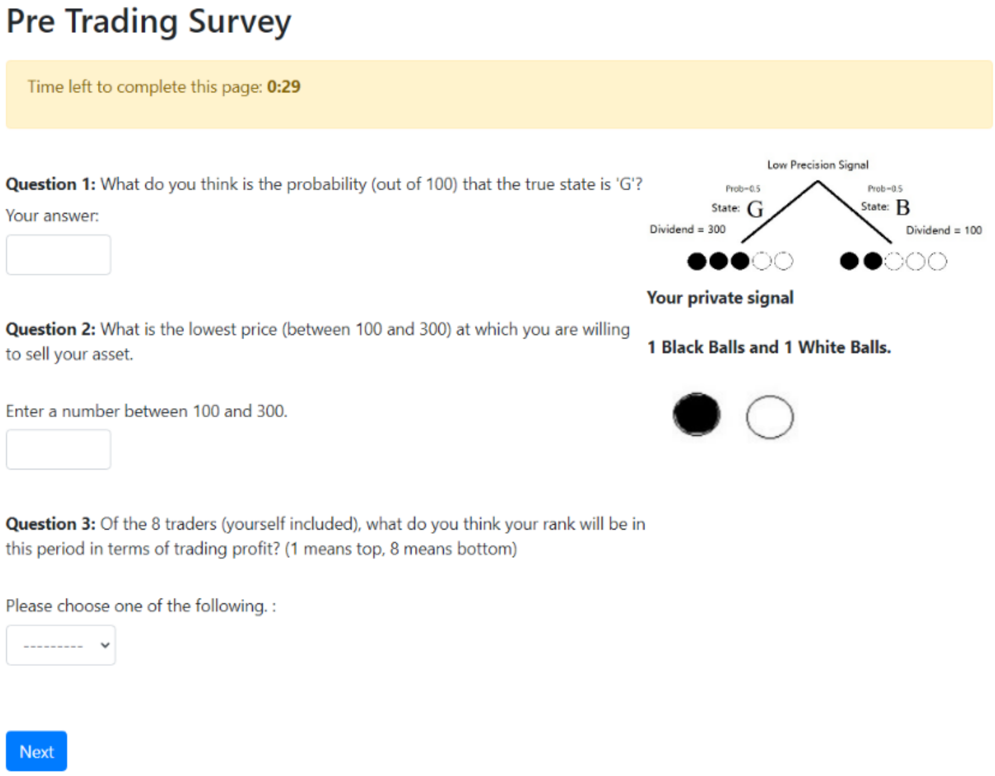
What do "low precision" and "high precision" mean? As shown in figure 1, H-sig has a higher proportion of black balls in the "good bag" and a higher proportion of white balls in the "bad bag". This means the H-sig has higher precision. For example, two black balls from the H-sig are a stronger indication that the true state is G than are two black balls from the L-sig. In half the periods of today's experiment, all traders get an L-sig and in the other periods all traders get an H-sig. You will see each period whether L-sig or H-sig is used.

Pre-Trading Survey

Before each trading period, you will be asked three questions as shown in Figure 2. You will get extra payment from doing this survey and the payment amount depends on your answers. To increase your payment, you should think it over carefully and report accurately **given your private signal** which is shown on the right-hand side of the page.

Question 1: What do you think is the probability (out of 100) that the true state is 'G'? Please enter a number that reflects your belief; if you think that, given your current private signal(s), it is 65% likely the true state is Good, then type 65 into the answer box.

Question 2: What is the lowest price (between 100 and 300) at which you are willing to sell your asset? You will soon have the opportunity to sell the asset


Pre Trading Survey
Time left to complete this page: 0:29
Question 1: What do you think is the probability (out of 100) that the true state is 'G'?
Your answer:
Low Precision Signal
Prob=0.5
State: G
Dividend = 300
Prob=0.5
State: B
Dividend = 100
Your private signal
1 Black Balls and 1 White Balls.
Question 2: What is the lowest price (between 100 and 300) at which you are willing to sell your asset.
Enter a number between 100 and 300.
Question 3: Of the 8 traders (yourself included), what do you think your rank will be in this period in terms of trading profit? (1 means top, 8 means bottom)
Please choose one of the following. :

Next

Figure 2

and must decide on what price to accept for each unit, given your beliefs about whether it will eventually pay 100 (if B) or 300 (if G). Please type your lowest acceptable price into the answer box for this question.

Question 3: Of the 8 traders (yourself included), what do you think your rank will be in this period in terms of trading profit? (1 means top, 8 means bottom) The computer will rank all traders' payoffs earned in the current trading from the highest (1st) to the lowest (8th). Please guess how your own trading profits will compare to other participants'. For example, if you guess that your trading profits will be the 2nd highest among the 8 traders, then select the number 2 and click "Next".

The payment for the survey is made in such a way that it is in your best interest to think carefully and respond truthfully to each question. If you are curious about the details of the payment, please see ¡Survey Question Procedure¿.

How to trade

Figure 3 shows the computer screen for trading. The right-hand side reminds you of your signal type and your private signal. There is a timer on the top counting down the time left. "Your Allocation" box tells you your current cash holding and number of asset units; it will update whenever you trade. You will start with two units of asset at the beginning of each trading period.

"Bids(Buy)" box shows the prices at which traders are offering to buy units of the asset and the "Asks(Sell)" box shows the prices at which sellers are offering to sell units. Bids are ordered from high to low since the highest bids are the most attractive offers to other traders. Likewise, asks are ordered from low to high since the lowest is the most attractive offer to other traders. If two bids/asks have the same price, the earlier offers will be displayed above the later ones. A bid/ask colored in green/blue indicates your own offer. “Trades” box in the middle shows the trading history in the market with the most recent trades on the top. Your own purchases/sales are colored in green/blue.

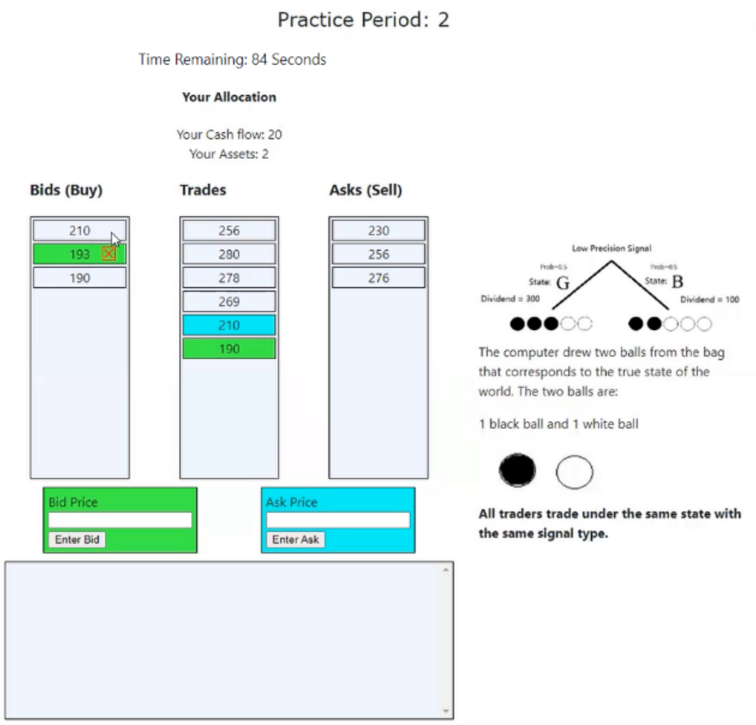


Figure 3

You submit a bid (buy offer) by typing in your price in the “Bid Price” box and clicking the Enter Bid button. Likewise, to offer to sell a unit, you type in the price in the “Ask Price” box and click Enter Ask. You can place at most 1 bid and 1 ask at the same time. If you enter a new bid or ask, it will replace your old offer. If your bid is higher than the best ask, then you will immediately buy a unit at that best ask. Similarly, entering an ask below the best bid held by someone else amounts to accepting that best bid. You can also directly accept someone else’s bid or ask by double clicking on it, and then clicking Accept in the pop-up window. You can buy at most 8 shares of assets.

It is more profitable to buy low and sell high. Check the trading window and see whether your Green buy prices seem low, and your Blue sell prices seem high compared to the other prices (in white) this period. The large "Error Message" box at the bottom of the trading screen will tell you if you made some kinds of mistakes, such as trying to sell when you currently hold zero units of the asset.

How do you get paid for trading in the market?

The payoff from trading in each period will be net cash flow (i.e., receipts from asset units sold minus cost of units bought), plus the payments on the assets you hold. For example, if you sold one unit for 280 and then you bought a unit for 260, your net cash flow would be 280 - 260 = 20. Moreover, since you started with 2 units of the asset, sold one unit and then bought one unit, your final asset holding is 2 units. Then your final payoff from trading in this period equals to 20+2*300=620 if the state turns out to be G, and 20+2*100=220 if the state is B.

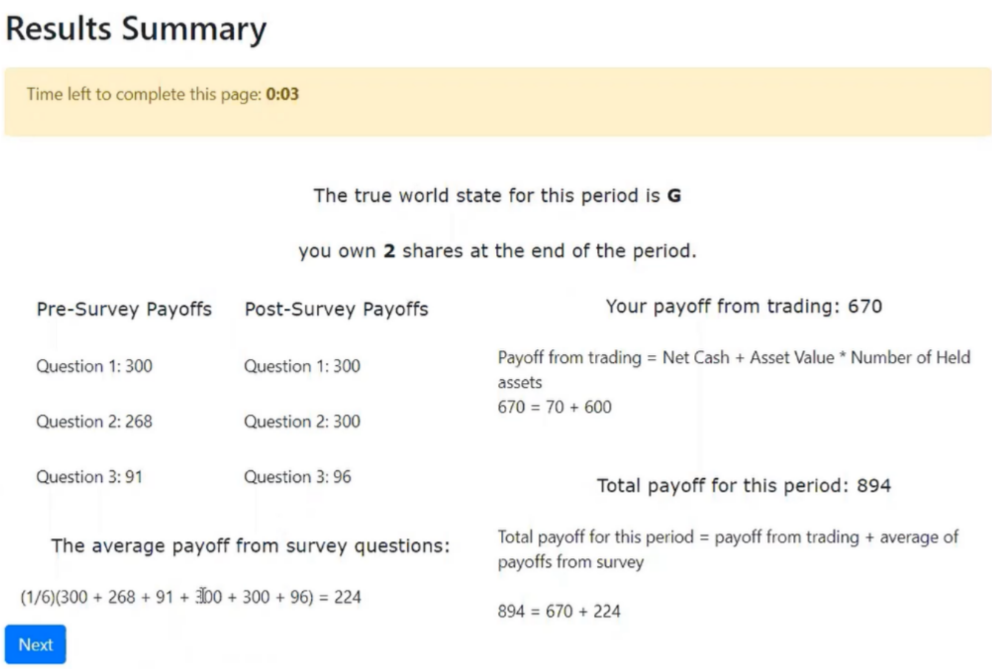
Post-Trading Survey

After each trading period, there will be another survey page that asks you the same three questions as in the pre-trading survey about your beliefs. The only difference between the post- and pre-trading survey is that the post-trading survey reminds you of the trading history from the current period, and that history might tell you something about what other traders' private signals are. So taking that history as well as your own private signal into account, you should **update your beliefs** before you answer the questions.

In addition, the trading history in this post-trading survey is ordered from high to low. The computer did this descending sorting for you to help you learn your trading strategy. If your buying prices (green) are, on average, lower than your selling prices (blue), then it means you are making profit. If it is the other way around, you need to adjust your trading strategy.

Feedback Page

After the post-trading survey page, you will see feedback pages similar to that shown below. These pages reveal the true state of the world and remind you of your final asset holding. They also show you the points you earned from answering survey questions and the payoff you earned from trading. They also show the total payoff for this period, which is "payoff from trading" + "average payoff from survey questions". After this feedback page, all traders go on together to the next period.



You will play 12 to 18 trading periods. Your final payoff for participating in today's session equals the sum of the total payoffs from all periods. Your final payoff is converted from experimental points to cash (US Dollars) at a rate announced at the beginning of the session, e.g., 1,000 experimental points = \$2.30 in cash.

C.2.3 Instructions for Other Treatments

All the instructions use the same format from the previous subsection and they only differ in the following points:

- At the end of the paragraph where we explain the difference between "low precision" and "high precision", we explain how the private signals are distributed among traders. This part distinguishes the treatments.
  - In the instructions of environment-a, we say "In half the periods of today's experiment, all traders get an L-sig and in the other periods all traders get an H-sig. You will see each period whether L-sig or H-sig is used."
  - In the instructions of environment-b, we say "In all the periods of today's experiment, half of the traders (by random) get an L-sig and the other half traders will get an H-sig. You will see each period whether you received a L-sig or a H-sig"
  - In environment-c, we say "In all the periods of today's experiment, you will know only your own signal and its precision. Other traders may have higher or lower precision than you, and their signals may differ — each trader gets fresh draws from either their own good bag or own bad bag, and those bags might have different numbers of black and white balls than your bags." We also explain that two signal precisions in the instructions are two examples.
- In the instructions of environment-c, figure 2 and figure 3 do not label the signal precision. Also, figure 3 needs to change the reminder message of how the private signals are distributed among traders accordingly below the private signal.

C.2.4 Survey Questions Procedure

Survey Questions Procedure

Question 1 (in both the pre and post trading survey): "What do you think is the probability (out of 100) that the true state is 'G'?"

How you will get paid: When you answer this question, you will be offered one unit of the asset. You can earn the dividend from this asset or exchange it with another asset that will be described below.

- Enter what you believe the probability is that the true state of the trading asset is "G". Let us denote this probability as R.
- We will offer you another asset called the N-asset, which is worth 600 with probability P_N and 400 with probability (1-P_N). (The probability P_N is determined randomly between 0 to 99.)
- If $R > P_N$, you keep your trading asset and earn the dividend from the asset. (300 for G and 100 for B).
- If $R \leq P_N$, you exchange your trading asset with us, thus hold the N-asset, and earn the dividend from the N-asset.

Question 2 (in both the pre and post trading survey): "What is the lowest price (between 100 and 300) at which you are willing to sell your asset."

How you will get paid: When you answer this question, you will be offered one unit of the asset. You can either earn the dividend from this asset or sell it back to us.

- You decide on the lowest price at which you are willing to sell your asset and enter that price into the computer.

- We will offer you a price to buy the asset back from you. The price is a random number between 100 and 300.

- If the price you are willing to sell is higher than our purchase price, you will keep your asset and earn that asset's value.

- If the price you are willing to sell is lower than our purchase price, you will sell your asset to us at the price we offered.

Question 3 (in both pre post trading surveys): "Of the 8 traders (yourself included), what do you think your rank will be in this period in terms of trading profit? (1 means top, 8 means bottom)"

How you will get paid: Your payoff from answering this question is determined by the formula $100 - (C - R)^2$. Where, C refers to the correct ranking of your payoff among all traders. R refers to your guess, which you entered into the computer.

Your best strategy for answering the three questions is to truthfully report!

Here is why (we use question 2 as an example to explain):

▷ You think this asset is worth at least 200 to you, so 200 is the lowest price you would willing to sell. This is your true belief. But you enter 250 into the box. You overstate your true belief.

▷ Then the computer will offer a price to buy, and that price is a random number between 100 and 300. Suppose the computer randomly chooses 240 to offer you for the asset.

▷ The computer cannot purchase the asset from you because 240 is lower than the lowest price you would accept to sell (250). So you keep your asset.

▷ This means you only earn 200; however, if you truthfully report, 200 into the computer, then the computer would purchase that asset from you at the price of 240, so you would earn 240.

▷ Your earning is higher when you truthfully report.

The payment methods for question 1 and question 3 are designed in the same way, which means there is no circumstance in which offering a probability/price/ranking not equal to your true belief is to your advantage; it can only decrease your earnings.

After you finish reading the written instruction, we will show you a video to explain, visually, what this game is about.

Bibliography

L. F. Ackert, B. K. Church, and P. Zhang. What affects the market's ability to adjust for optimistic forecast bias? evidence from experimental asset markets. *Journal of economic behavior & organization*, 66(2):358–372, 2008.

A. A. Alchian. Uncertainty, evolution, and economic theory. *Journal of political economy*, 58(3):211–221, 1950.

A. Alesina, A. Miano, and S. Stantcheva. The polarization of reality. In *AEA Papers and Proceedings*, volume 110, pages 324–28, 2020.

J. R. Anderson. *The adaptive character of thought.* Psychology Press, 2013.

N. H. Anderson. *Foundations of information integration theory*, volume 578. Academic Press New York, 1981.

N. Aragón and R. P. Roulund. Confidence and decision-making in experimental asset markets. *Journal of Economic Behavior & Organization*, 178:688–718, 2020.

J. S. Banks. Price-conveyed information versus observed insider behavior: A note on rational expectations convergence. *Journal of Political Economy*, 93(4):807–815, 1985.

N. Barberis, A. Shleifer, and R. Vishny. A model of investor sentiment. *Journal of financial economics*, 49(3):307–343, 1998.

K. Barron. Belief updating: does the ‘good-news, bad-news’ asymmetry extend to purely financial domains? *Experimental Economics*, 24(1):31–58, 2021.

K. Barron, A. Becker, and S. Huck. Motivated political reasoning: The emergence of belief-value constellations. *Unpublished paper*, 2022.

O. E. Barron and H. Qu. Information asymmetry and the ex ante impact of public disclosure quality on price efficiency and the cost of capital: Evidence from a laboratory market. *The Accounting Review*, 89(4):1269–1297, 2014.

G. M. Becker, M. H. DeGroot, and J. Marschak. Measuring utility by a single-response sequential method. *Behavioral science*, 9(3):226–232, 1964.

B. Biais, D. Hilton, K. Mazurier, and S. Pouget. Judgemental overconfidence, self-monitoring, and trading performance in an experimental financial market. *The Review of economic studies*, 72(2):287–312, 2005.

K. Bregu. Overconfidence and (over) trading: The effect of feedback on trading behavior. *Journal of Behavioral and Experimental Economics*, 88:101598, 2020.

J. Caballé and J. Sákovics. Speculating against an overconfident market. *Journal of Financial Markets*, 6(2):199–225, 2003.

C. Camerer and K. Weigelt. Information mirages in experimental asset markets. *Journal of Business*, pages 463–493, 1991.

C. F. Camerer and T.-H. Ho. Violations of the betweenness axiom and nonlinearity in probability. *Journal of risk and uncertainty*, 8(2):167–196, 1994.

G. Charness and C. Dave. Confirmation bias with motivated beliefs. *Games and Economic Behavior*, 104:1–23, 2017.

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.

L. Choo, T. R. Kaplan, et al. Information aggregation in arrow–debreu markets: an experiment. *Experimental Economics*, 22(3):625–652, 2019.

J. Clark and L. Friesen. Overconfidence in forecasts of own performance: An experimental study. *The Economic Journal*, 119(534):229–251, 2009.

T. E. Copeland and D. Friedman. Partial revelation of information in experimental asset markets. *The Journal of Finance*, 46(1):265–295, 1991.

A. Coutts. Good news and bad news are still news: Experimental evidence on belief updating. *Experimental Economics*, 22(2):369–395, 2019.

P. Crosetto, A. Filippin, P. Katuščák, and J. Smith. Central tendency bias in belief elicitation. *Journal of Economic Psychology*, 78:102273, 2020.

R. Croson and J. Sundali. The gambler's fallacy and the hot hand: Empirical data from casinos. *Journal of risk and uncertainty*, 30(3):195–209, 2005.

C. Cueva and I. Iturbe-Ormaetxe. Motivated beliefs about stock returns. *Available at SSRN 3905142*, 2021.

R. Deaves, E. Lüders, and G. Y. Luo. An experimental test of the impact of overconfidence and gender on trading activity. *Review of finance*, 13(3):555–575, 2009.

D. Eil and J. M. Rao. The good news-bad news effect: asymmetric processing of objective information about yourself. *American Economic Journal: Microeconomics*, 3(2):114–38, 2011.

N. Epley and T. Gilovich. The mechanics of motivated reasoning. *Journal of Economic perspectives*, 30(3):133–40, 2016.

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

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

L. G. Epstein, J. Noor, and A. Sandroni. Non-bayesian learning. *The BE Journal of Theoretical Economics*, 10(1), 2010.

S. Ertac. Does self-relevance affect information processing? experimental evidence on the response to performance and non-performance feedback. *Journal of Economic Behavior & Organization*, 80(3):532–545, 2011.

T. Q. Fan, Y. Liang, and C. Peng. The inference-forecast gap in belief updating. *Unpubished Manuscript*, 2021.

G. Fellner-Röhling and S. Krügel. Judgmental overconfidence and trading activity. *Journal of Economic Behavior & Organization*, 107:827–842, 2014.

R. Forsythe and R. Lundholm. Information aggregation in an experimental market. *Econometrica: Journal of the Econometric Society*, pages 309–347, 1990.

D. Friedman and M. Aoki. Inefficient information aggregation as a source of asset price bubbles. *Bulletin of Economic Research*, 44(4):251–279, 1992.

D. Friedman and J. Rust. The double auction market. *Addison-Wesley*, 1993.

D. Friedman, G. W. Harrison, and J. W. Salmon. The informational efficiency of experimental asset markets. *Journal of political Economy*, 92(3):349–408, 1984.

M. Glaser and M. Weber. Overconfidence and trading volume. *The Geneva Risk and Insurance Review*, 32(1):1–36, 2007.

M. Glaser, T. Langer, and M. Weber. True overconfidence in interval estimates: Evidence based on a new measure of miscalibration. *Journal of Behavioral Decision Making*, 26(5):405–417, 2013.

K. Gödker, P. Jiao, and P. Smeets. Investor memory. *Available at SSRN 3348315*, 2021.

R. Gonzalez and G. Wu. On the shape of the probability weighting function. *Cognitive psychology*, 38(1):129–166, 1999.

A. Gotthard-Real. Desirability and information processing: An experimental study. *Economics Letters*, 152:96–99, 2017.

B. Greiner. Subject pool recruitment procedures: organizing experiments with orsee. *Journal of the Economic Science Association*, 1(1):114–125, 2015.

D. M. Grether. Recent psychological studies of behavior under uncertainty. *The American Economic Review*, 68(2):70–74, 1978.

D. M. Grether. Bayes rule as a descriptive model: The representativeness heuristic. *The Quarterly journal of economics*, 95(3):537–557, 1980.

D. M. Grether. Testing bayes rule and the representativeness heuristic: Some experimental evidence. *Journal of Economic Behavior & Organization*, 17(1):31–57, 1992.

Z. Grossman and D. Owens. An unlucky feeling: Overconfidence and noisy feedback. *Journal of Economic Behavior & Organization*, 84(2):510–524, 2012.

S. A. Heger and N. W. Papageorge. We should totally open a restaurant: How optimism and overconfidence affect beliefs. *Journal of Economic Psychology*, 67:177–190, 2018.

D. Hilton, I. Regner, L. Cabantous, L. Charalambides, and S. Vautier. Do positive illusions predict overconfidence in judgment? a test using interval production and probability evaluation measures of miscalibration. *Journal of Behavioral Decision Making*, 24(2):117–139, 2011.

J. C. Hoffart, S. Olschewski, and J. Rieskamp. Reaching for the star ratings: A bayesian-inspired account of how people use consumer ratings. *Journal of Economic Psychology*, 72:99–116, 2019.

M. Hoffman and S. V. Burks. Worker overconfidence: Field evidence and implications for employee turnover and returns from training. Technical report, National Bureau of Economic Research, 2017.

C. A. Holt and A. M. Smith. An update on bayesian updating. *Journal of Economic Behavior & Organization*, 69(2):125–134, 2009.

D. Huffman, C. Raymond, and J. Shvets. Persistent overconfidence and biased memory: Evidence from managers. *Pittsburgh: University of Pittsburgh*, 2019.

T. Ida and R. Okui. Can information alleviate overconfidence? a randomized experiment on financial market predictions. *A Randomized Experiment on Financial Market Predictions (April 3, 2020)*, 2020.

D. Kahneman and A. Tversky. Subjective probability: A judgment of representativeness. *Cognitive psychology*, 3(3):430–454, 1972.

D. Kahneman and A. Tversky. Prospect theory: An analysis of decision under risk. In *Handbook of the fundamentals of financial decision making: Part I*, pages 99–127. World Scientific, 2013.

E. Karni. A mechanism for eliciting probabilities. *Econometrica*, 77(2):603–606, 2009.

E. Kirchler and B. Maciejovsky. Simultaneous over-and underconfidence: Evidence from experimental asset markets. *Journal of Risk and Uncertainty*, 25(1):65–85, 2002.

C. Kraemer, M. Nöth, and M. Weber. Information aggregation with costly information and random ordering: Experimental evidence. *Journal of Economic Behavior & Organization*, 59(3):423–432, 2006.

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

C. M. Kuhnen. Asymmetric learning from financial information. *The Journal of Finance*, 70(5):2029–2062, 2015.

R. Lunawat. Learning from trading activity in laboratory security markets with higher-order uncertainty. *Journal of Behavioral and Experimental Economics*, 90:101611, 2021.

J. Magnani and R. Oprea. Why do people violate no-trade theorems? a diagnostic test. Technical report, Working paper, 2017.

M. Marinacci. Learning from ambiguous urns. *Statistical Papers*, 43(1):143–151, 2002.

D. W. Massaro and D. Friedman. Models of integration given multiple sources of information. *Psychological review*, 97(2):225, 1990.

C. Meier and L. De Mello. Investor overconfidence in experimental asset markets across market states. *Journal of Behavioral Finance*, 21(4):369–384, 2020.

J. Michailova and U. Schmidt. Overconfidence and bubbles in experimental asset markets. *Journal of Behavioral Finance*, 17(3):280–292, 2016.

J. B. Miller and A. Sanjurjo. Is it a fallacy to believe in the hot hand in the nba three-point contest? *European Economic Review*, 138:103771, 2021.

M. M. Mobius, M. Niederle, P. Niehaus, and T. S. Rosenblat. Managing self-confidence:

Theory and experimental evidence. Technical report, National Bureau of Economic Research, 2011.

D. A. Moore and P. J. Healy. The trouble with overconfidence. *Psychological review*, 115(2):502, 2008.

C. N. Noussair, C. R. Plott, and R. G. Riezman. An experimental investigation of the patterns of international trade. *The American Economic Review*, 85(3):462–491, 1995. ISSN 00028282. URL http://www.jstor.org/stable/2118183.

C. N. Noussair, C. R. Plott, and R. G. Riezman. An experimental investigation of the patterns of international trade. In *International Trade Agreements and Political Economy*, pages 299–328. World Scientific, 2013.

T. Odean. Volume, volatility, price, and profit when all traders are above average. *The journal of finance*, 53(6):1887–1934, 1998.

R. Oprea and S. Yuksel. Social exchange of motivated beliefs. *Journal of the European Economic Association*, 2020.

P. Ortoleva. Modeling the change of paradigm: Non-bayesian reactions to unexpected news. *American Economic Review*, 102(6):2410–36, 2012.

E. Oster, I. Shoulson, and E. Dorsey. Optimal expectations and limited medical testing: evidence from huntington disease. *American Economic Review*, 103(2):804–30, 2013.

L. Page and C. Siemroth. How much information is incorporated into financial asset prices? experimental evidence. *The Review of Financial Studies*, 34(9):4412–4449, 2021.

G. Pennycook and D. G. Rand. Examining false beliefs about voter fraud in the wake of the 2020 presidential election. *The Harvard Kennedy School Misinformation Review*, 2021.

C. R. Plott and S. Sunder. Efficiency of experimental security markets with insider information: An application of rational-expectations models. *Journal of political economy*, 90(4):663–698, 1982.

C. R. Plott and S. Sunder. Rational expectations and the aggregation of diverse information in laboratory security markets. *Econometrica: Journal of the Econometric Society*, pages 1085–1118, 1988.

D. Prelec et al. The probability weighting function. *ECONOMETRICA-EVANSTON ILL-*, 66:497–528, 1998.

M. Rabin. Inference by believers in the law of small numbers. *The Quarterly Journal of Economics*, 117(3):775–816, 2002.

M. Rabin and J. L. Schrag. First impressions matter: A model of confirmatory bias. *The quarterly journal of economics*, 114(1):37–82, 1999.

S. Sunder. Market for information: Experimental evidence. *Econometrica: Journal of the Econometric Society*, pages 667–695, 1992.

M. Thaler. The supply of motivated beliefs. *arXiv preprint arXiv:2111.06062*, 2021.

A. Tversky and D. Kahneman. Belief in the law of small numbers. *Psychological bulletin*, 76(2):105, 1971.

A. Tversky and D. Kahneman. Judgment under uncertainty: Heuristics and biases. *science*, 185(4157):1124–1131, 1974.

A. Tversky and D. Kahneman. Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–291, 1979.

A. Tversky and D. Kahneman. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and uncertainty*, 5(4):297–323, 1992.

Z. Wang. Motivated beliefs, markets, and information aggregation. *Markets, and Information Aggregation (April 8, 2022), Unpublished Manuscript*, 2022.

Z. Wang, D. Friedman, and T. Bowen. Overconfidence and market performance. *Available at SSRN 4308529*, 2022.

C. Wilde and W. Li. Belief formation and belief updating under ambiguity: Evidence from experiments. *Unpublished paper*, 2019.

G. Wu and R. Gonzalez. Curvature of the probability weighting function. *Management science*, 42(12):1676–1690, 1996.

H. Zhang. Evolutionary justifications for non-bayesian beliefs. *Economics Letters*, 121 (2):198–201, 2013.

C. Zhao. Representativeness and similarity. Technical report, Working Paper, Princeton University, 2016.