Will Bayesian Markets Induce Truth-telling? —An Experimental Study

Yan Xu *

University of Vienna

This Version: October, 2020

Abstract

Abstract: This paper tests the performance of the Bayesian market (Baillon (2017)), a new mechanism that incentivizes truth-telling from crowds. A participant in a Bayesian market trades a belief asset whose value is determined by other participants trading positions. In the truth-telling Bayesian Nash equilibrium, a participant will reveal her private signal through buying or short-selling an asset when she believes others are also truthful. I create three Bayesian markets in the lab, varying in the belief uncertainties regarding participants truthfulness in the market. I find Bayesian markets effective when participants reasonably believe others are truthful. They are less effective when participants beliefs are subject to noise and updating biases. A further investigation of participants bids and ask prices demonstrates how bubbles are formed and impede the performance of Bayesian markets. With belief uncertainty, participants exhibit under-inference bias or ignorance in processing their private signals. The speculators in the market further amplify the updating bias, creating a speculative trend about the belief asset. This finding provides new insights into the role of belief uncertainty and updating biases in the evolvement of market bubbles.

JEL classification:
Keywords: Truth-telling mechanism, Bayesian market, asset bubbles, belief bias.

*Contact: yan.xu@univie.ac.at.

1 Introduction

Information elicited from dispersed individuals is increasingly crucial to many knowledge-gathering and decision-making tasks. Researchers are conducting extensive social-economic surveys to collect knowledge about human perceptions. Crowd-sourcing platforms are keen to solicit informative answers from online communities. Customers are steering their purchases to products with reliable customer reviews. In most cases, information providers are compensated for their time and efforts, if at all, not for their truthfulness; consequently, they may provide uninformative or untruthful answers. A worker who labels images on MTurk[1], for instance, may type random tags given the wage is fixed. A survey-taker may lie about sensitive questions involving drug abuse or impaired driving. When the answers from the crowd are unreliable, the algorithms and decisions based on them will be biased. How to encourage respondents to report their private information truthfully is the biggest challenge in information elicitation.

When a verifiable truth is present to condition payment, many mechanisms create incentives for truth-telling. Consider the proper scoring rules (Winkler (1969)).[2] A respondent submits a probabilistic forecast of an event, and then the report is graded against the objective truth – the realization of the event or its frequency – by a score. Typically, the closer a report is to the underlying truth, the higher the score it will attain, and correspondingly the higher the reward will be (see Gneiting and Raftery (2007) and Gneiting and Katzfuss (2014)). The prediction market is another popular elicitation mechanism.[3] A participant buys or sells a contract paying one dollar if an event occurs. Wolfers and Zitzewitz (2006) showed that asset prices on prediction markets are close to the mean belief of all participants. Hence, an agent whose prediction probability of the event is higher than the average belief will buy a contract. An agent whose prediction is lower than the average belief will sell a contract.

Observing the objective truth is crucial for scoring rules and prediction markets in incentivizing truth-telling. In the example of scoring rules, the event outcome or its actual frequency serves as an evaluation gauge for all possible forecasts from individuals. In the presidential prediction market, the election outcome verifies profits for all contracts. Both mechanisms condition individuals' payoffs with the observable outcome of the underlying event. The payment schemes guarantee that truthful reports of their opinions and beliefs of

[1]MTurk (https://www.mturk.com) is a crowd-sourcing platform enabling individuals to perform Human Intelligence Tasks to earn money.

[2]Early work also include Brier (1950), Good (1950), Winkler and Murphy (1968), and Savage (1971). Schlag et al. (2015) compared various scoring rules. Offerman et al. (2009) generalized them for non-expected utilities. Schotter and Trevino (2014) reviewed their empirical implementations.

[3]Early inspirations of prediction market are from Hayek (1945) and Fama (1970)'s discussions on the efficient information aggregation in financial markets. Related reviews include Manski (2006), Tziralis and Tatsiopoulos (2007), Berg et al. (2008a), and Berg et al. (2008b). Recent applications can be found in Arrow et al. (2008), Dreber et al. (2015), Camerer et al. (2016), and Camerer et al. (2018). In practice, the Iowa Electronic Markets (https://iemweb.biz.uiowa.edu/) are running examples of collecting individual opinions on political elections through market transactions.

the event yield the highest expected payoff. Principally, the validity of an elicitation mechanism hinges upon how well the monetary incentives are aligned to informants' truthful revelations. When the underlying truth is subjective or costly to verify, it is less intuitive to apply the incentive alignment principle. For questions like "Have you ever engaged in questionable research practice?", "Do you believe computers will outsmart humans?" or "Are you happy?", there is no natural benchmark against which respondents' answers are evaluated, and truthful ones are rewarded. Nevertheless, subjective truths like feelings, judgments, and emotions are prominent in modern life.

Miller et al. (2005)'s peer prediction method and Prelec (2004)'s Bayesian truth serum (BTS) are two classes of mechanisms eliciting the unverifiable truth.[4] They exploit the relationship of private information among the population and construct benchmarks for respondents' answers based on peers' answers. In the example regarding the survey question, "Are you happy?", the implementer of the peer prediction mechanism (called a center) is assumed to know the prior distribution of private information. She transforms each respondent's answer into a belief about a reference's answer through Bayesian updating and further evaluates the belief through a proper scoring rule. BTS relaxes the assumption of the center knowing the prior distribution. Each participant answers"yes" or "no" to the question and also predicts the proportion of participants in a large population answering "yes." The center assigns each participant a prediction score and an information score. The prediction score induces truthful prediction of the distribution of signal reports. The information score exploits the implied Bayesian reasoning about population frequencies and rewards truthful signal reports that are more common than collectively predicted ones. Since peers' answers are verifiable, truth-telling proves to be a Bayesian Nash Equilibrium for both mechanisms.

The validity of peer predictions and Bayesian truth serums in practical implementations remains an open question. Gao et al. (2014) experimentally tested the peer prediction method on MTurk. They found players were more likely to coordinate at uninformative equilibria, suggesting a failure of peer prediction in inducing the truth-telling equilibrium. John et al. (2012) employed BTS incentive schemes and surveyed 2000 psychologists on their involvement in questionable research practices. They found BTS induced a higher self-admission rate. Weaver and Prelec (2013) tested BTS in a recognition questionnaire containing foil brand names or scientific terms. They showed that participants claimed to recognize fewer foils in BTS groups than in control groups, further supporting BTS's capability in inducing truth-telling. In general, it is very difficult to explain to subjects how these score-based mechanisms incentivize truthful answers. Indeed, subjects in these experiments could not link their actions to payoffs directly, and they were suggested to believe that truth-telling was in their best interests. Suspicions and demand effects may arise. Shaw et al. (2011) employed BTS as a contextual manipulation on MTurk and found that workers performed

[4] Jurca and Faltings (2006, 2009) extended the peer prediction method in avoiding uninformative equilibria. Parkes and Witkowski (2012) proposed Robust BTS for small population. Radanovic and Faltings (2013, 2014) generalized RBTS to non-binary and continuous signals.

significantly better, even though they were not financially rewarded by BTS. They argued that out-performance of BTS might be attributed to confusion and cognitive demand.

Baillon (2017) proposed a new institution, called Bayesian market, to simplify the practical implementation of BTS. Through a market where private information is linked to asset transactions, individuals are rewarded for the truthful revelation of subjective truth. In particular, a Bayesian market associated with the question "Are you happy?" works as follows: First, each agent has an opportunity to participate in the market; he can buy (sell) at most one asset by submitting a "yes" ("no") report. Then, the market price of the asset is randomly drawn, and agents are asked whether they would like to trade at the price. The asset value is the realized proportion of people answering "yes" to this question. Similar to the information score in BTS, agents who are truly happy expect higher proportion of "yes" report and thus have higher valuation for a share of asset. Hence, they are more likely to buy an asset than those who feel unhappy. Under mild assumptions, Bayesian markets predict a truth-telling BNE.

This paper aims to test the validity of Bayesian markets in inducing truth-telling. Although Bayesian markets theoretically reward truth-telling, they may not necessarily induce it in practice. One obstacle is the possible confusion and cognitive demands. Another one inherits in the definition of truth-telling BNE: the truthful report is optimal for an agent if he believes all other agents are truthful. Moreover, the belief disturbances in others' truthfulness might be the main source of cognitive demands. To understand how belief disturbances influence the effectiveness of Bayesian markets, I manipulate individual beliefs over others' truthfulness in Bayesian markets. By varying the extent of disturbances, this paper answers the following three questions: (1) Will Bayesian markets induce the best response of truth-telling from individuals when they believe all others are truthful? (2) Will Bayesian markets induce the best response of truth-telling when participants expect that some agents may lie? (3) Will Bayesian markets induce the truth-telling BNE?

I answer these questions by constructing three types of laboratory Bayesian markets, each featuring a setting where participants' belief of others' truthfulness is distorted to some degree. I create links of private information among agents through bingo cages. They imply common priors and further generate private signals for all agents. After drawing a signal, each agent updates beliefs, determines the trading position, and submits a bid or ask in the Bayesian market. The distortion of beliefs is achieved by Algorithm Agents (AAs), who always report their private signals and correct beliefs in asset value. By varying proportions of AAs among eight agents in Bayesian markets, I expose human agents (HAs) to different degrees of belief in truthfulness. I consider three treatments: 1HA, 3HA, and 8HA treatment. Each one aims to address one research question proposed before.

I find Bayesian markets effectively induce truthful reports of private signals and posterior expectations of asset value when the belief system is perfect. When there are disturbances of agents' belief in others' truthfulness, Bayesian markets are less effective. These treatment effects are be explained by the arise of bubbles in the market. Due to noises in belief, agents under-inference private signals and are more likely to buy than short sell assets, making the expectation of asset value higher than the fundamental value. Furthermore, this belief is confirmed by the realization of asset value, and thus bubbles arise.

The rest of the paper is structured as follows. In section 2, I briefly describe Bayesian market mechanism and its theoretical predictions. Section 3 introduces the experimental design and procedures. In section 4 and **??**, I analyze the performances of Bayesian markets and subjects' beliefs and choices. Section 6 concludes this paper and discusses potential future works.

2 Bayesian market mechanism

2.1 Model setup

Consider the simplest case that n homogeneous risk neutral agents are surveyed with the same binary question — "Are you happy?". Population n is assumed to be infinite for theoretical simplicity in this section but will be relaxed in the experiment. Agent i's private signal (his subjective truth) is assumed to be a random variable $T_i \in \{Y, N\}$. Before receiving a realization of the private signal, agent i believes that signals are drawn from a joint distribution $f(T_1, T_2, ..., T_n)$. This prior distribution is assumed to be a common knowledge among all agents. In the case of binary signals, a common prior can also be described by $f(\omega)$, where ω is the proportion of agents whose private signal is Y. Given a signal realization denoted as t_i, agent i updates his belief of ω through $f(\omega \mid T_i = t_i)$.

A Bayesian market is set up to elicit individuals' private signals. Each participant can trade at most one asset by submitting a yes/no report to the question of interest. The value of the asset v is the proportion of agents reporting yes. The market price of the asset, denoted as p, is randomly drawn from a commonly known uniform interval $[0, 1]$. Individuals' reports determine the trading positions on the market. After knowing the price, each agent is asked whether he would like to buy an asset at p if he reported "yes", and whether he would like to sell an asset at p if he reported "no". To prevent agents from learning private information from each other through the direct trade, there is a market maker ("she") between buyers and sellers. She executes a transaction for a buyer (seller) if a majority of the agents reporting "yes" ("no") is willing to buy (sell).

Since agents cannot influence the market price, their trading decisions rely on how they evaluate the asset after knowing their private signals. To better understand how people's posterior expectations of the asset value are influenced by disturbances in belief systems, I deviate from Baillon (2017) and implement the BDM method (Becker et al. (1964)) to elicit people's posterior beliefs. The trading decision is replaced by submitting a bid/ask price:

A bid price $b_i \in [0,1]$ represents the highest price to buy an asset and an ask price $a_i \in [0,1]$ is the lowest price to sell an asset. Instead of calculating complex scores under alternative mechanisms, agents on Bayesian markets undertake asset transactions that are more familiar for them and thus may create greater engagement. The market maker collects all bids and asks from participants and calculates the average bid and ask price as her buying ($\bar{a}$) and selling ($\bar{b}$) price. The trading rule resembles the majority rule. Specifically, a trade occurs for buyer i if $b_i \geq p \geq \bar{a}$, under which both the buyer and the market maker are willing to trade at price p. Similarly, a seller will successfully short sell an asset to the market maker if $a_i \leq p \leq \bar{b}$.

After a market closes, the market maker will liquidate each asset at its settlement value v. Trading buyers will receive an amount of money equaling v, and trading sellers need to pay back an asset at a cost of v. Hence, profit is $v - p$ for a trading buyer and $p - v$ for a trading seller. Those who fail to trade receive zero.

2.2 Theoretical predictions of Bayesian markets

I summarize the theoretical predictions on how agents trade in Bayesian markets and how the market transactions result in truth-telling equilibrium in this subsection. Formal proofs can be found in Baillon (2017) and Baillon (2016).

A Bayesian market's validity in truth-inducing relies on the link between an agent's private signal and his expectation of others' signals. Before receiving private signals, an agent i forms an expectation of Y-type in population according to the common prior $f(\omega)$. After receiving a private signal t_i, agent i updates the expectation based on $f(\omega \mid t_i)$. Signals are assumed to be "impersonally informative". By "informative", the signals provide information about population frequency ω; by "impersonal," agents who receive the same signals will learn in the same way of ω. Mathematically speaking, $f(\omega \mid t_i) = f(\omega \mid t_j) \Leftrightarrow t_i = t_j$ implies both "informative" and "impersonal" property. Condition $f(\omega \mid t_i) = f(\omega \mid t_j) \Rightarrow t_i = t_j$ means that different types of agents will formulate different posterior belief. On the other hand, condition $f(\omega \mid t_i) = f(\omega \mid t_j) \Leftarrow t_i = t_j$ requires that same types of respondents will draw same inference of ω.

The signal structure and the implied Bayesian reasoning form the basis for incentive alignment in Bayesian markets. When receiving different private signals, agents formulate different expectations of others' private information. In particular, agents who are truly happy (Y-type) will expect more happy people in the population than those who are not (N-type). Namely, $E(\omega \mid t_Y) > E(\omega \mid t_N)$. Notice that it is a relative comparison. Both types of agents may expect a minority of N-type in population, however, by exploiting the information in their true answers, happy agents still expect that Y-type signals are more common in the population than what unhappy agents expect.

Prediction 1. *Agent who participates in a Bayesian market will submit his posterior expectation of asset value $E(v \mid t_i)$ as his bid or ask price.*

After deciding trade positions on the market, buyers/sellers choose bid/ask prices to maximize posterior expected payoff and the optimal one will be $E(v \mid t_i)$. In particular, contingent on agent i's buy/sell decision, individuals' willingness to pay for an asset is $b_i = E(v \mid t_i)$ and their willingness to accept an asset is $a_i = E(v \mid t_i)$. This prediction is not surprising due to the implementation of BDM method. When market prices are randomly determined, it is incentive compatible for an agent to report his underlying willingness to pay/accept, which equals to the posterior expectation of asset value. For notation simplicity, I denote $E(v \mid t_Y) \equiv \omega_Y$ and $E(v \mid t_N) \equiv \omega_N$ as the posterior expectation for a Y-type and a N-type agent, respectively. If all agents follow the optimal bid and ask strategy, the market maker's buying price will be $\bar{b} = \omega_Y$ and her selling price will be $\bar{a} = \omega_N$.

Prediction 2. *Truth-telling is a BNE in Bayesian markets.*

According to the definition of BNE, when all other agents truthfully report their private signals, the asset value is the same as the frequency of the Y-type signal (ω). By Prediction 1, active agents in the market will report their posterior expectations of asset value, producing ω_Y and ω_N as a market maker's buying price and selling price, respectively. A Y-type agent expects a higher asset value than an N-type agent and thus is willing to buy an asset. Similarly, an N-type agent is willing to sell an asset. In equilibrium, both types of agents will participate in markets and truthfully reveal their private signals through their buy/sell decisions.

3 Experimental design and procedures

3.1 Market structure

Theoretical analyses reveal that the validity of Bayesian markets in inducing truth-telling relies on two critical assumptions – common prior and "impersonally informative" signals. Both are directly defined on the signal spaces. Unlike mechanisms like peer prediction, Bayesian markets do not require the formulation structure of private signals. In practice, it yields convenience. For example, I may don't know the states of the world and how they affect individuals' underlying happiness. However, in some other scenarios, such as writing customer reviews, agents form opinions after a realization of the state of the world. Since this paper focuses on whether and how Bayesian markets induce truth-telling, I design the experiment in accordance with the case where both common priors and signals are determined by states of the world.

There are two states of the world, represented by two types of bingo cages in the experiment. Both types of cages contain 100 balls and act as possible random devices to generate private signals for all participants in the market. To avoid cognitive differences in private information, I use more neutral labels of signals – red ball and blue ball, corresponding to Y-type and N-type in model setup. Since agents share the same belief about states of the world and about how signals are generated for each possible state before receiving private signals, they share a common prior $f(\omega)$, where ω in our experiment is the proportion of red ball in population. Once a bingo cage is chosen as a random device, the state of the world will be realized, and then agents will receive a ball generated by the chosen cage. Hence, signals are both "impersonal" and "informative": agents who receive different balls expect different distributions of ω, and those who receive the same balls update belief in ω in the same way.

In the example of a typical Bayesian market illustrated in Figure 1, two typeA bingo cages contain 67 out of 100 red balls, and two other cages contain 33. Since one from four bingo cages will be randomly chosen to generate private signals, the common prior of red ball proportion, denoted as ω, is:

$$f(\omega) = \begin{cases} \frac{1}{2} & \text{if } \omega = 0.67 \\ \frac{1}{2} & \text{if } \omega = 0.33 \\ 0 & \text{otherwise} \end{cases}$$

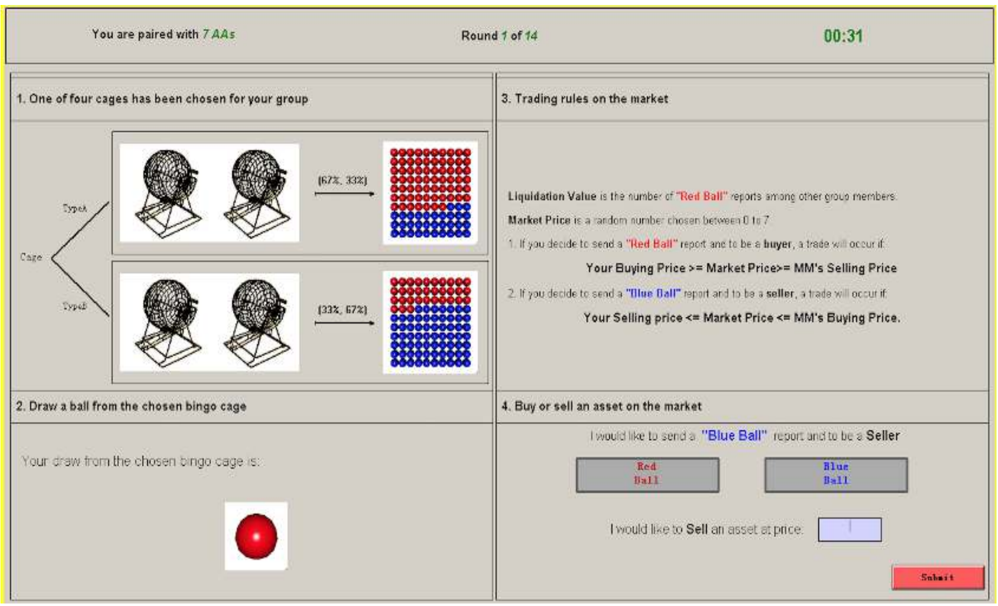

Figure 1: Decision screen

When the state of the world is realized, that is, a bingo cage is chosen, an agent can click a button and draw a ball, showing in the left bottom area of the screen. Even though he does not know which cage has been chosen, he can update the likelihood of each state and further infer the posterior probability of ω in the market. For instance, after receiving a red ball, an agent expects a 67% chance with which the chosen bingo cage is of typeA. Based on that, he further updates the belief of ω through:

$$f(\omega|t_i = Y) = \begin{cases} 0.67 & \text{if } \omega = 0.67 \\ 0.33 & \text{if } \omega = 0.33 \\ 0 & \text{otherwise} \end{cases}$$

Based on this posterior distribution, his expectation of ω is $E(\omega|t_i = Y) = 0.56$

Given common priors and private signals, agents can buy/sell an asset by submitting a red ball/blue ball report, and a bid/ask price in Bayesian markets. This can be done in the right bottom area of the interface. I allow for at most eight agents in the market and rescale asset value, bid/ask prices, and market price from $[0, 1]$ to $[0, 7]$. By Proposition 4, Bayesian markets work for any $n \geq 4$. The choice of $n = 8$ is a compromise between market interaction and experiment control. On the one hand, it generates multiple realizations of asset value; on the other hand, it still guarantees control in the lab. The asset value is the number of red ball reports submitted by all opponent agents, and the market price is randomly drawn from a truncated normal distribution on $(0, 7)$. Other distributions also work in Bayesian markets as long as they are commonly known. However, by weighting intermediate prices with higher probabilities, I try to increase the chance of transactions between traders and market makers in the experiment.

I design the experiment in a repeated setting – each session consists of 14 periods, and each period is a new Bayesian market with a different prior. Appendix I lists all sets of parameters that determine common priors and private signals. Although the mechanism is described as a one-shot game, it is difficult for subjects to immediately recognize the best response and the equilibrium of the game. Hence, I introduce learning as a possible de-bias device and focus on testing whether Bayesian markets will induce truth-telling in convergence. To facilitate learning, I keep the group members fixed during the whole session and provide each subject full information of signals, decisions, and profits after the current period completes. An example of review screens is depicted in Figure 2.

It is important to emphasize that Bayesian markets reward truth-telling even when ground truth is not accessible. In our experiment, the truth, which is the ball generated by the chosen bingo cage, is verifiable. Potential relaxation of this procedure will be discussed later.

3.2 Experimental treatments

According to Proposition 3, Bayesian markets predict a truthful-telling BNE among all participants – it is in a participant's best interest to report his private information if he believes all other agents are truthful. However, there is no guarantee that subjects will


You decided to Sell an asset at Price 3.00
The Liquidation Value for you is: 3.00
The Market Price of the asset is: 3.05
MM's Buying Price is 3.71 and MM's Selling Price is 2.86

Seller's Trading condition: MM's Buying Price >= Market Price >= Your Selling Price is satisfied.
(3.71 >= 3.05 >= 3.00)
Your payoff in this round = 0.05+ 3.00 = 3.05 tokens

Here are decisions and profits for all participants in your group:

Player	Draw	Report	Buyer/Seller	Buy/Sell Price	Liquidation Value	Trade or Not	Profit
You	Blue Ball	Blue Ball	Seller	3.00	3.00	Trade	0.05
Agent1	Blue Ball	Blue Ball	Seller	2.86	3.00	Trade	0.05
Agent2	Blue Ball	Blue Ball	Seller	2.86	3.00	Trade	0.05
Agent3	Red Ball	Red Ball	Buyer	3.71	2.00	Trade	-1.05
Agent4	Red Ball	Red Ball	Buyer	3.71	2.00	Trade	-1.05
Agent5	Blue Ball	Blue Ball	Seller	2.86	3.00	Trade	0.05
Agent6	Blue Ball	Blue Ball	Seller	2.86	3.00	Trade	0.05
Agent7	Red Ball	Red Ball	Buyer	3.71	2.00	Trade	-1.05

Next Round

Figure 2: Review screen

hold such beliefs. For example, it's reasonable for agents to expect that some others might be confused by the mechanism or might prefer lying. This strict belief requirement is the most arresting obstacle that prevents subjects from telling the truth. To understand whether and how Bayesian markets perform, I control subjects' beliefs in the experiment.

I construct the market with two types of agents – Human Agent (HA) and Algorithm Agent (AA). HAs are human participants, whose buy/sell decisions and bid/ask prices are what I am interested in. On the other hand, AAs are programmed to report private signals and posterior expectations of asset value truthfully. To make a participant believe that all other agents are truth-telling, I can group him with AAs. A participant's belief of other agents' truthfulness can be controlled by the number of HAs and AAs in the market. Following this simple idea, I design three treatments that differ in the degree of others' truthfulness.

The first treatment, called 1HA treatment, aims to test whether Bayesian markets will induce the best response of truth-telling from individuals when their opponents are restricted to be truthful. Each human participant in this treatment will be grouped with 7 AAs. The task of an agent is to decide whether to buy or to sell an asset and to submit a bid or ask price of the asset. Since AAs are programmed to tell the truth, human agents will reasonably believe that all other agents are truthful. Based on Proposition 3, it is optimal for a human subject to truthfully report his type and his posterior expectation of asset value in this treatment.

In the second treatment, called 3HA treatment, I allow for disturbances of HA's belief to test whether the best response of truth-telling from individuals is robust under Bayesian markets. Each HA in the experiment will be grouped with two other HAs and 5 AAs. AAs are still truthful, but HAs may expect some other human agents to lie. However, it is still optimal for a HA to report truthfully under strict participation condition[5] and if he expects a red ball type agent are more likely to report red ball than a blue ball type agent.[6] I design group composition and prior parameters to make sure that this condition is satisfied as much as possible. The simulation of 10000 draws shows that in 90% cases, it indeed is satisfied.

The third treatment, 8HA treatment, further tests whether Bayesian markets will select out the truth-telling equilibrium in a more realistic setting where there is no restriction imposed on belief. Each human participant will be grouped with seven other HAs. They may form different beliefs in other HAs' truthfulness. However, they are rewarded for formulating correct belief and further coordinating at the truth-telling BNE in Bayesian markets.

3.3 Experimental procedures

I ran our experiment in May and June 2016 at Erasmus University of Rotterdam. A total of 87 subjects were recruited in 4 sessions, and each session lasted around 90 minutes. The average payment was 21.80 euros. The number of subjects, groups and average payoffs for each treatment are described in Table 1:

	1HA	3HA	8HA
Subjects	25	30	32
Groups	25	10	4
Observations	350	420	448
Payoffs	22.78	21.22	21.57

Table 1: Summary of three experimental treatments

Upon arrival, each subject was randomly assigned an ID and was guided to a computer desk. Then they were asked to read instructions[7] and finish a quiz regarding trade and profits in Bayesian markets. After that, all subjects would trade in markets for 14 periods. Each period is a new Bayesian market with different common priors and private information. At the end of the experiment, I also asked subjects to fill out a non-incentivized questionnaire regarding their understanding of the experiment, their socio-demographic characteristics, and self-reported risk attitudes.

The monetary unit in the experiment is called tokens, each worth 0.5 euro. After a market had closed in one period, asset price was randomly chosen from (0,7). Then a market

[5]Strict participation means that an agent who expects strictly positive payoff will participate in the market. This condition is satisfied in truth-telling equilibrium.

[6]This result can be found in Baillon (2017)

[7]The instruction for 1HA treatment is attached in Appendix II.

maker calculated the average bid/ask price and determined whether a trade would occur for each agent. She would also liquidate all assets in the market and calculate participants' profits. Since agents might lose money, they were endowed with three tokens at the beginning of each period. The total payment for each subject was the sum of endowments and profits in all 14 rounds. Subjects in 3HA and 8HA treatment faced similar interfaces and decision tasks. The only difference was the status bar showing different group compositions in the decision screen. To better understand how agents behave on Bayesian markets under different belief systems, I kept three treatments comparable in all aspects: parameter settings for common priors were the same; each set of parameters appeared in the same order; prices were determined by same distributions.

4 Experimental results

4.1 Aggregate truthful rate

The disturbances in people's belief systems do influence the validity of Bayesian markets. At the aggregate level, the proportion of truthful reports varies with the treatment stimulus. Table 2 summarizes the average percentages of truthful reports in three treatments. In 1HA treatment, 80% of the reports submitted by agents are their private signals. And this number is 68% for 3HA and 63% for 8HA Treatment, respectively.

	1HA	3HA	8HA
Truthful Rate	0.80	0.68	0.63
Observations	350	420	448

Table 2: Aggregate rate of truthful reports in three treatments

Unsurprisingly, all three truthful rates are lower than the theoretical value of 100%. Even though the experiment satisfies prior and information assumptions of Bayesian markets, there are other implied structural assumptions imposed on agents for the prediction that truth-telling is a best response in 1HA and 3HA treatment and a BNE in 8HA treatment. For instance, subjects are required to use sophisticated Bayesian reasoning to predict others' signals in the same way. Given the noise inherent in real settings, the truthful rate in 1HA treatment provides reasonable supports for the validity of Bayesian markets.

Pairwise comparisons between aggregate truthful rates in different belief settings reveal valuable information about the performance of Bayesian markets. In 1HA treatment, 80% of the reports submitted by agents are the same as private signals. But this number is much lower in 3HA and 8HA treatment. Mann-Witney tests show that the truthful rate in 1HA treatment is significantly higher than that in 3HA and 8HA treatment, but there is no significant difference between 3HA and 8HA. These results imply that when there are more HAs in the market, participants feel uncertain about others' truthfulness, and Bayesian markets are less effective in inducing truth-telling.

4.2 Individual truthful rate

The treatment differences in truthful rates remain at the individual level. Figure 3 depicts the frequency of truth-telling for each subject in three treatments. In 1HA treatment, 25% of subjects report their private signals truthfully, and 21% only lie once during 14 periods. However, in 3HA and 8HA treatment, few subjects are fully truthful, and as a consequence, the aggregate truthful rates are lower than that in 1HA treatment. In terms of inducing truth-telling from individuals, again, I find Bayesian markets outperformed in 1HA treatment.

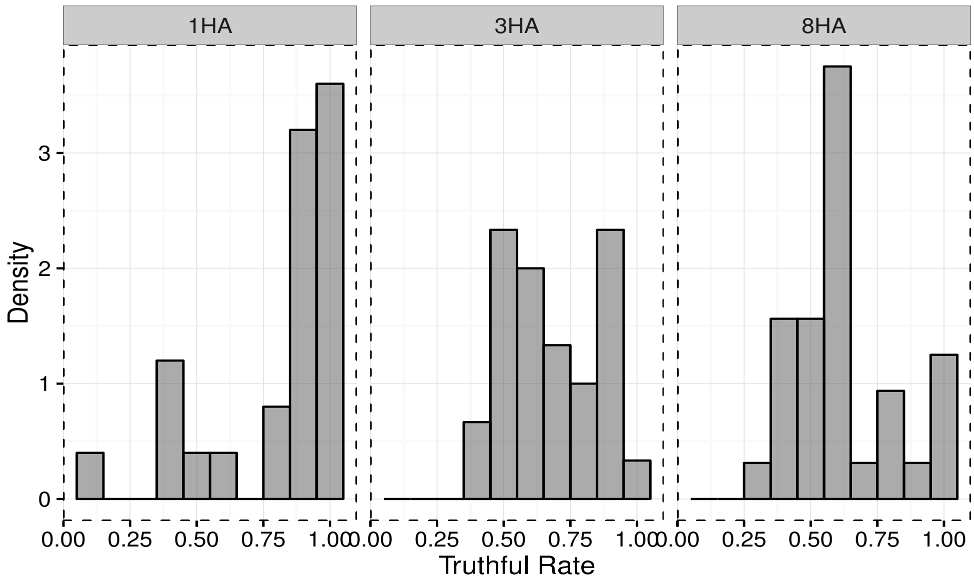

Figure 3: Histogram of individual truthful rate in three treatments

A natural question is: what causes treatment effects of both aggregate and individual truthful rates? Since the truthful rate captures the average level rather than a dynamic change of truth-telling, heterogeneity in strategy evolution among treatments might explain. Another possible source for treatment effects is the heterogeneity in signal effects. By signal effects, I define it as the change of truth-telling incentives due to differences in signals. When the number of HAs increases, there will be more noise in the market, triggering different signal effects between treatments. In the following subsections, I will check these possible sources separately.

4.3 Heterogeneity in strategy evolution

One possible source of treatment effects is the heterogeneity in strategy evolution among treatments. For instance, truthful rates in three treatments may start at a similar level, but evolve at different speeds and therefore result in different aggregate levels. Since cognitive sophistication required by each experiment is increasing with the number of HAs in the market, I expected learning speed to be the highest in 1HA treatment, medium in 3HA and the lowest in 8HA treatment.

Figure 4 illustrates the time series of average truthful rates in three treatments. The solid line is for 1HA treatment, dotted for 3HA, and dashed for 8HA treatment. There is little evolution trend in both 1HA and 8HA treatment. For 1HA treatment, the truthful rate is persistently high, limiting space for further improvement. For 8HA treatment, the interaction among human agents introduces too much noise for effective learning. However, there is a prominent learning trend in 3HA treatment, where the truthful rate started at the lowest level of 57% and increased to 80%, close to that in 1HA treatment.

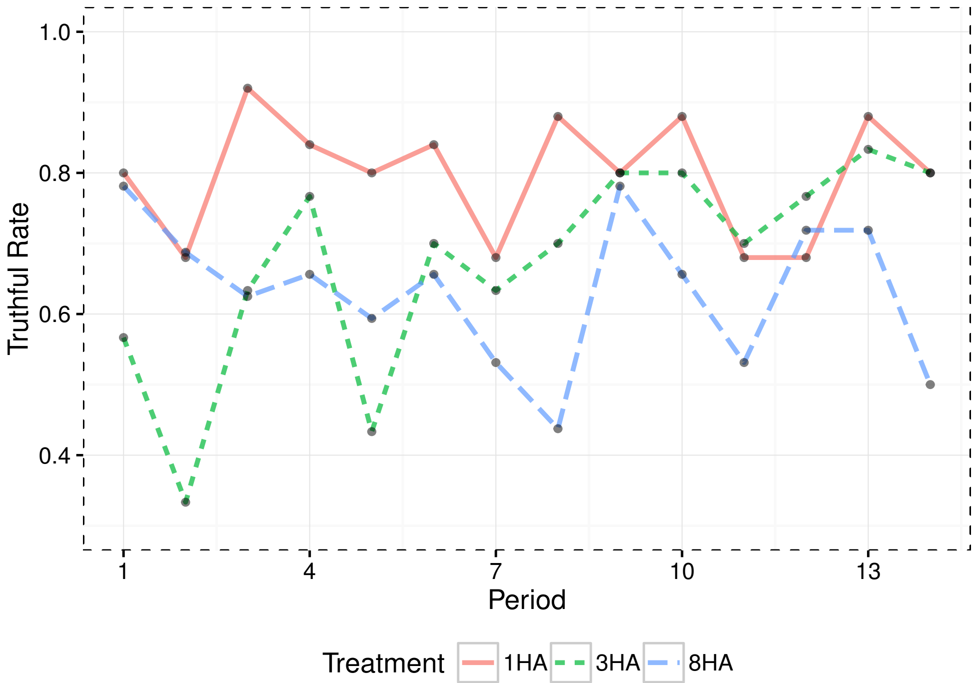

Figure 4: Truthful rate by periods

I do find heterogeneity in strategy evolution among treatments, but it is between 3HA and 1HA (8HA) treatment. Even though their truthful rates were similar, Bayesian markets with 3HA might outperform those with 8HA in the long run because of the learning effect. However, strategy evolution fails to explain treatment effects between 1HA and 8HA treatment.

4.4 Heterogeneity in signal effects

Another source of treatment effects is the heterogeneity in signal effects, meaning agents respond to private signals in different ways. I calculate the posterior truthful rates conditional on private signals in Table 3. Given a private signal is a red ball, more than 80% of reports are truthful, and it holds for all three treatments. However, after receiving a blue ball signal, agents are more likely to lie: 40% reports submitted by agents in 3HA are untruthful, and this number increases to 49% in 8HA treatment. Mann-Witney tests between two conditional truthful rates show significant differences in 3HA and 8HA treatment, but not in 1HA treatment.

	1HA	3HA	8HA
Pr(Truthful | Sig = Red)	0.85	0.81	0.85
Pr(Truthful | Sig = Blue)	0.77	0.60	0.51

Table 3: Truthful rate conditional on private signals

At first sight, this result is quite puzzling because different labels of private signals should not affect truthful rates in a systematic way. Moreover, I design the experiment to be symmetric of signals: first, it is equally likely to receive a red ball or a blue ball on average; second, each market in the experiment corresponds to another market where labels of two signals are swapped. Therefore, posterior truthful rates should be equal for both types of signals.

However, inherit elicitation procedures of Bayesian markets may lead to different perceptions of different signals. Specifically, signal reports submitted by agents are associated with their buy/sell decisions. Implied transaction positions of private signals, rather than labels, may affect incentives of truth-telling. Since these incentives are triggered by market conditions, they are highly likely to vary with group compositions for each treatment. A plausible explanation is that agents are more likely to buy than short sell assets because they are more experienced with purchase decisions. This phenomenon is quite normal in market institutions, and I call it "buying inclination". In the experiment, as the number of HAs increased, markets become noisier. The choice of buying/selling is more complicated, and thus there is a more severe buying inclination.

I further test buying inclination and its implication for the truthful rate in Table 4. The benchmark of the buying rate is around 0.36 – the realized frequency of red ball signal. It implies that if all reports are truthful, 36% of them should decide to buy assets. However, buying rates calculated from agents' decisions in three treatments are all higher than benchmarks. In particular, buying rates in 3HA and 8HA treatment are as high as 55% and 63%, implying more severe bias towards buying than in 1HA treatment.

The third and fourth row of Table 4 report the truthful rates conditional on buyer/seller decisions. Among all reports from sellers, the majority of them are truthful. However, among all buyer reports in the market, there is a significant difference between 1HA and 3HA (8HA) treatment. 50% reports submitted by buyers are untruthful in 8HA treatment, compared with 53% in 3HA and 67% in 1HA treatment.

	1HA	3HA	8HA
Benchmark	0.36	0.36	0.37
Buying Rate	0.45	0.55	0.63
Pr(Truthful | Buyer)	0.67	0.53	0.50
Pr(Truthful | Seller)	0.90	0.85	0.86

Table 4: Buying rate and truthful rate

The aggregate truthful rate is an average of truthful rates of buyers and sellers with the buying rate as a weight. Taken together, I conclude that treatment effects are driven by signal effects. More specifically, two factors jointly played a role: first, agents in 3HA and 8HA treatment are more likely to buy than to sell assets; second, buyers are more likely to lie than sellers.

5 Explaining treatment effect: bid and ask prices

In this section, I will explain the differences in truthful rates in three treatments via agents' bid and ask prices for the asset in Bayesian markets.

5.1 Aggregate bid and ask prices

Figure 5 demonstrates the time series of the submitted price, the theoretical price, and the value of the asset in Bayesian markets. The solid line is the average price submitted by subjects in each period, representing an average HAs' ex-ante valuations of the asset. The dashed line shows the correct predictions based on their signals. And the dotted line is the ex-post asset value.

I focus on the ex-ante and ex-post prediction gaps in each treatment. The first one means the difference between the submitted and the theoretical prices, and the second one is the difference between the submitted prices and the realizations of asset value. First, the ex-ante prediction gaps in all treatments are positive for all fourteen periods, meaning that an average HA in a Bayesian market predicts a higher asset value than a AA does. Therefore, HAs are more likely to buy assets than AAs in all treatments, consistent with the buying inclination shown in the previous section.

There is a significant difference between 1HA (3HA) and 8HA treatment in terms of the

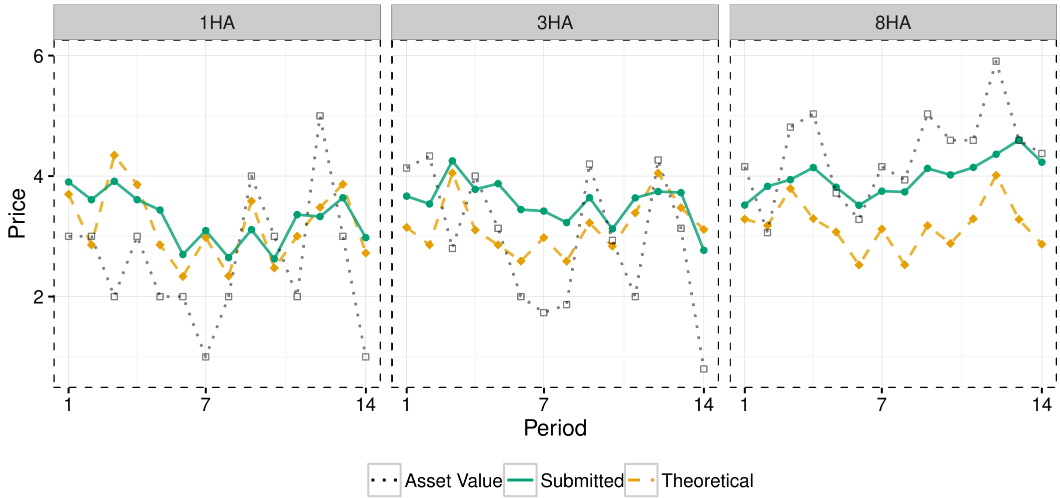

Figure 5: Aggregate bid/ask price and the theoretical price

size and trend of the ex-ante prediction gaps. In both 1HA and 3HA treatments, ex-ante prediction gaps are quite small and exhibit no trend. In other words, an average HA in these two treatments correctly predicts the realization of asset value. While in 8HA treatment, the prediction gap starts at a substantial level and is increasing over time, indicating that HAs are persistently over-predicting asset value. There is also a significant difference between 1HA (3HA) and 8HA in their ex-post prediction gaps. However, the direction is opposite to the ex-ante gaps. An average HA in 8HA treatment predicts the asset value more accurately than the agent in 1HA and 3HA treatment does.

The treatment effects of both ex-ante and ex-post prediction gaps are consistent with each other and jointly explain the trend-chasing of submitted prices in 8HA treatment. Since subjects in 8HA treatment over-predicts asset value more severely, they are over-buying assets in the markets. This further drives up the ex-post realization of the asset value. In other words, their beliefs of the asset are self-confirmed by their decisions, leading to bubbles in Bayesian markets.

How does the trend of ex-ante and ex-post prediction gaps relate to the truthful rate? Figure 6 shows the same time series of price gaps for truthful and untruthful reports. It confirms that untruthful subjects are the main drive for the over-prediction of asset value and the resulting over-buying in markets. On average, truthful agents submit prices close to theoretical ones. They act as Bayesian agents who exploit private information to update belief. Untruthful agents, on the other hand, severely over-predicts asset value. They act as trend-chasers who take into account possible bubbles in the market. For 8HA treatment, comparisons between truthful and untruthful agents are more evident: truthful subjects submit predictions reflecting the fundamental value of the asset and untruthful subjects chase bubbles and submit predictions reflecting the market valuation of the asset.

Figure 6: Aggregate bid/ask price and the theoretical price by truthfulness

Since untruthful reports are in the form of either receiving a red ball but reporting blue ("RB") or receiving a blue ball but reporting red ("BR"). I further check the time series of the submitted price, theoretical prices, and asset value based on private signals and buy/sell decisions in 8HA treatment in Figure **??** of Appendix III. The patterns are similar. Agents who receive blue ball signals and decide to be buyers are trend-chasers. They submit prices close to the realization of asset values. Agents who receive red ball signals and choose to be sellers act as Bayesians. Their prices reflect that they formulate correct beliefs in asset value after knowing their private information.

Combining with the buying inclination in the previous section, I find that agents in 8HA treatment who receive blue ball signals have higher valuations of the asset than its fundamental value. They thus are more likely to buy assets, further raising the realized asset value. The asset value indeed confirms their belief, and they would continue to chase the trend and finally cause bubbles in the market.

Another point is that the time series of prices are aggregated with four Bayesian markets, each corresponding to a group of 8HAs in the treatment. Appendix III shows the same time series of submitted price, theoretical price, and asset value for each group. The aggregate ex-ante prediction gaps are mainly driven by Group 24. In all three markets, there is no clear evidence of trend-chasing and the resulting market bubbles. Deleting Group 24 will reduce the extend of bubbles in 8HA treatment.

5.2 Shirking, speculation, and updating bias

A Bayesian agent will update his private signal and then submits his expected number of red reports as a bid (or ask) price. When he believes that all others are truthful, his bid/ask price is the expected proportion of red balls among seven agents $7E(\omega \mid t_i)$ (Proposition 2 and 3). In the experiment, ω takes two values, ω_A and ω_B. Each one represents the proportion of red balls in the corresponding types of the cage. Therefore, the bid/ask price should be bounded by two values of 7ω. Moreover, subjects' posterior beliefs about the state A (the chosen cage is of type A), denoted as $\pi(A \mid t_i)$ can be backed out by their bid/ask prices through $c_i = \pi(A \mid t_i)\omega_A + (1 - \pi(A \mid t_i))\omega_B$.

The bid/ask prices reveal agents' updating and participation decisions in a Bayesian market. Different patterns may explain how belief environment affects the validity of Bayesian markets. For instance, if an agent submits a bid lower than $7\omega_B$ or an ask higher than $7\omega_A$, he cannot trade an asset whatever the asset price is, and others' decisions are. In other words, agents in a Bayesian market can avoid trading an asset by submitting low bids or high asks on purpose. I call such reports being "shirking". On the other hand, If an agent submits a bid higher than $7\omega_A$ or an ask lower than $7\omega_B$, he will guarantee himself a trade of the asset. I call such reports being "speculative". When the bid/ask price is within the range of $7\omega_B$ and $7\omega_A$, the agent's posterior belief of the state is rationalized by a regular probability on $[0, 1]$, I call such reports being "Bayesian".

Table 5 reports the proportion of each types of bid/ask patterns. Around 10% of bid/ask prices show that agents prefer no participation in the market. Agents in 3HA and 8HA treatment are significantly more likely to be speculative. In 8HA treatment, most of the speculative reports are buyers who submit very low bid prices to make sure successful trades. This is consistent with the trend-chasers in the market. The majority of subjects valuations of the asset are still justified by Bayesian reasoning, and agents in 1HA treatment are significantly more Bayesian.

Does the difference in speculative and Bayesian patterns between 1HA and 3HA (8HA) treatment explain the treatment effects in truthful rate. After excluding shirking agents, the treatment effects remain. After further excluding speculative agents, the size of the treatment effects is smaller. The result implies that the disturbances in people's beliefs over others' truthfulness induce updating biases. Moreover, the speculative agents in the

	1HA	3HA	8HA
#Obs	350	420	448
Shirking	9%	10%	9%
Speculative	15%	22%**	29%**
Bayesian	75%	67%**	62%**
#Bayesian	265	283	279
Note:			**p<0.05;

Table 5: Proportion of different types of bid/ask price patterns.

market further enlarge the biases. The interaction between updating biases and speculation motives raises the bubbles in the market and impedes the truth-telling in Bayesian markets.

I focus on the bid/ask prices in the category of "Bayesian" and further examine people's updating biases in three treatments. Are the implied posteriors consistent with Bayesian updating? If not, do they support lying in the transaction of an asset? According to Bayes' Theorem of belief updating, we have the following formula of the posterior odds after receiving a signal of red ball:

$$\frac{\pi(A \mid Red)}{\pi(B \mid Red)} = \frac{p(A)}{p(B)} \times \frac{p(Red \mid A)}{p(Red \mid B)}.$$

Taking logs to both sides of the equality, I estimate the updating biases associated with the prior information and the diagnostic information of private signals separately by the following Grether (1980) regression:

$$\ln \frac{\pi_{ij}(A \mid Red)}{\pi_{ij}(B \mid Red)} = \beta_0 + \beta_1 \ln \frac{p_j(A)}{p_j(B)} + \beta_2 \ln \frac{p_j(Red \mid A)}{p_j(Red \mid B)} + \epsilon_{ij}.$$

The dependent variable is subject i's posterior odds after receiving a red ball in the period j. The independent variables include the prior odds and the likelihood ratio for the same problem j. When the subject is a perfect Bayesian, the coefficients for both prior odds and likelihood ratios are equal to one. Following the literature (see Benjamin (2019)), there are four updating biases. The coefficient of prior odds determines whether the subject has a bias of base-rate neglect ($\hat{\beta}_1 < 1$) or confirmation bias ($\hat{\beta}_1 > 1$). The coefficient of likelihood ratio determines whether the subject has under-inference ($\hat{\beta}_2 < 1$) or over-inference ($\hat{\beta}_2 > 1$) bias.

Table 6 reports the regression results. The coefficients of the log of prior odds are close to those in the literature. Subjects generally exhibit base-rate neglect. A joint F-test of $\hat{\beta}_1$ shows that the updating bias of base-rate neglect is similar across three treatments. The coefficients of the log of the likelihood ratio (diagnostic information) show that agents under-infer from the private signal. In 1HA treatment, the magnitude of the under-inference

(0.77) is significant and comparable to those estimated in standard Bayesian tasks. However, estimates of $\hat{\beta}_2$ are not significant, indicating that agents in 3HA and 8HA treatment ignore the private information. A joint F-test of $\hat{\beta}_2$ shows that the under-inference biases are significantly different between 1HA and 3HA (8HA) treatment. The difference in updating biases is consistent with the prediction gaps in Bayesian markets. When there are speculative trends, agents chase the trend and ignore their private signals. In 3HA treatment, AAs will stabilize the trend, while in 8HA treatment, the trend persists and evolves market bubbles.

	Dependent: LogPostOdd		
	1HA	3HA	8HA
Constant ($\hat{\beta}_0$)	-0.08	-0.09	0.41***
	(0.13)	(0.12)	(0.14)
LogPriorOdd ($\hat{\beta}_1$)	0.33*	0.53***	0.37**
	(0.19)	(0.09)	(0.15)
LogDiagOdd ($\hat{\beta}_2$)	0.77**	0.13	0.15
	(0.38)	(0.09)	(0.12)
Observations	265	283	279
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 6: Estimating the belief updating biases for three treatments

5.3 Heuristics

5.3.1 Keep and switch heuristics

A simple heuristic of learning is keeping or switching strategies conditional on profits in the previous period. Under this heuristic, subjects keep the strategy if it yields positive profit in the last round, and they switch to alternative strategies if otherwise. Figure 7 captures the frequencies of the switch and keep patterns for each treatment. There are four types of switch and keep pattern: "Keep truth-telling", "Keep lying," "Truth-telling to lying," and "Lying to truth-telling," each corresponding to the dotted, the solid, the dot-dashed and the dashed line. The truth-telling is quite focal in all treatments. The frequency of "Keep truth-telling" stays steady around 70% in 1HA treatment but is more volatile in 3HA and 8HA treatment. Remarkably, there is an increasing trend in 3HA for the pattern of "Keep truth-telling", which explains the learning effect of truthful rate in 3HA treatment. All the other three patterns are much volatile.

The change of frequencies of switch patterns may rely on previous profits. Table 7 lists the conditional switch rates for each treatment. I find more switches in both 3HA treatment and 8HA treatment. However, conditional on previous profits, there is no significant difference between switch frequencies. Collectively, when there are more noises in oth-

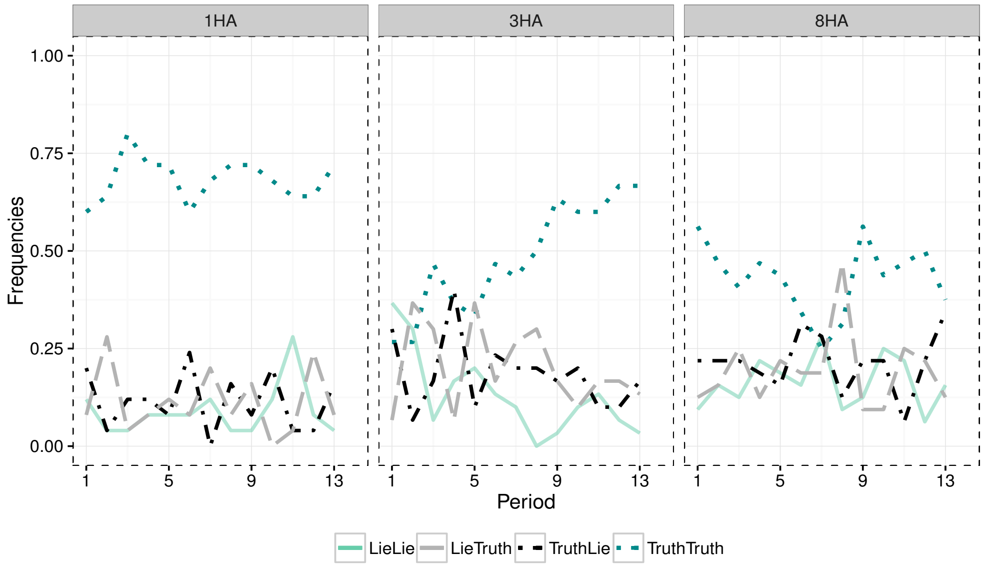

Figure 7: Time series of strategy keep and switch

ers' truthfulness, subjects try alternative strategies, switching from truth-telling to lying or from lying to truth-telling. However, these switches do not depend on previous profit.

Previous profit	Switch strategy	1HA	3HA	8HA
Positive	truth to lie	0.12	0.19	0.16
	lie to truth	0.04	0.21	0.28
Negative	truth to lie	0.11	0.19	0.25
	lie to truth	0.16	0.21	0.15
No Trade	truth to lie	0.11	0.18	0.22
	lie to truth	0.15	0.20	0.18

Table 7: Switch patterns conditional on previous profit

5.3.2 Imitation heuristic

Another simple heuristic is to imitate strategies that yielded the highest profit in the previous rounds. In the experiment, when each period completed, subjects learned all information about the last period on a review screen. They may check strategies and the corresponding profits of others and further adjust his strategy in the next round. Table 8 reports the imitation rate and winner's truthful rate for each treatment. Imitation rate is calculated as the frequency of which a subject's current strategy is the same as the winner's strategy in the previous round.

First, winner's truthful rate winner's truthful rates are close to one in 1HA and 3HA treatment, implying truth-telling yielded the highest profit on average. This result is not surprising because of the existence of AAs in the market. AAs are always truthful and will be

Treatment	1HA	3HA	8HA
Imitation rate	0.80	0.67	0.55
Winner's truthful rate	0.99	0.96	0.69

Table 8: Imitation rates in three treatments

winners in the equilibrium. When there are no AAs to stabilize the market price, bubbles arise, and lying can be the winning strategy. Hence the truthful rate for winners in 8HA treatment is significantly lower.

Second, the imitation rate in 1HA treatment is 80%, significantly higher than that in 3HA and 8HA treatment. Notice that the imitation rate here captures coincidence, rather than causality between subjects' strategy and winners' strategy. Since it is impossible to isolate imitation from other heuristic or strategic considerations solely from revealed decisions, I cannot conclude that subjects in 1HA are more likely to imitate the winners' strategy. But the imitate rates show a higher correlation between subjects' strategy and the winner's strategy in 1HA treatment.

How does the imitation heuristic affect truth-telling in each treatment? Figure 8 depicts the scatter plot of the imitation rate and the truthful rate. The correlation is quite strong in 1HA and 3HA treatment. But in 8HA treatment, a high imitation rate may not imply (or be implied by) high truthful rate because the winning strategy is less likely to be truthful.

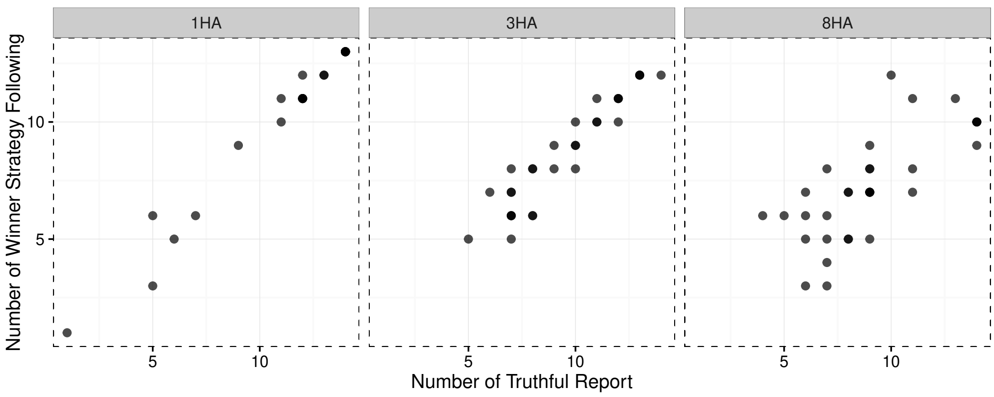

Figure 8: Relationship between imitation rate and truthful rate

6 Conclusion and Discussion

This paper tests the validity of Bayesian markets. Will they induce truth-telling, either as a best response or as equilibrium from individuals. I construct Bayesian markets in an experiment and manipulate human agents' beliefs in other players' truthfulness in three settings. I find Bayesian markets effectively induce truth-telling as a best response when agents reasonably believe that all their opponents are also truthful. In particular, agents submit truthful reports of private signals in most cases (80%) and form correct posterior expectations of asset value. However, when agents suspect that some of their opponents lie, Bayesian markets are less effective. Participants in the 3HA and 8HA treatment are less likely to reveal private information, and they over-predict asset value.

The raise of bubbles in 3HA and 8HA treatments explains how the belief noises affect the performance of Bayesian markets. Even though the majority of subjects' posterior beliefs after knowing their private signal are rationalizable by Bayesian reasoning, they reveal more severe under-inference when there are noises in their belief systems. Subjects in 3HA and 8HA treatment ignore their private signals and focus on the trends of asset realizations. I interpret the process of bubbles as follows: (1) Common priors and impersonal private information are not sufficient to induce common posterior expectations. (2) Due to the speculative buyers in the market, traders who are supposed to sell assets predict higher-than-fundamental asset value and thus are more likely to buy assets. (3) over-buying in the market raised the realization value of assets, which further confirmed ex-ante belief. (4) more traders will ignore the private signal and chase the trend to buy assets. AAs in 3HA treatment manage to stabilize the speculative trend. While in 8HA treatment, bubbles raise in the Bayesian markets.

A closer check of what generated the bubble in the first place revealed buying inclination among participants, which also explained a relatively lower frequency of truth-telling in 3HA and 8HA treatment. To be specific, when traders are more uncertain about the behaviors of other traders, they would be more likely to buy than short sell assets simply because of the familiarity of the former context. The speculative buyers reflected in the bid/ask patterns also confirm the buying inclination. Initially, it might just raise the average expectation of asset value slightly, but it had the potential to trigger self-confirming belief in the market and to bring about bubbles. A critical lesson for the practical implementation of the Bayesian market mechanism is that I may improve data quality by familiarizing participants with the concept of short-sell.

One concern for the validity of Bayesian markets is the seemingly forced participation in our experiments. In the truth-telling equilibrium, Bayesian markets predict that all agents, regardless of their signals, choose to participate in the market since truth-telling yields a strictly positive payoff in expectation. However, when an agent believes that the market is out of equilibrium, he may prefer to opt-out. For instance, even though an agent with a blue ball signal recognizes that there are bubbles on the market, he is forced to "ride the bubble" because otherwise, he may lose. It should be noted that traders can opt-out in our experiment by submitting a bid of 0 or an ask of 7. However, few subjects did so (2.5% for buyers and 1.8% for sellers). The analyses of bid/ask patterns show that the out-out rates are not the drive for the difference in truthful rate in three treatments. Many subjects may not realize there is an option to step out of the market. By introducing treatment with a salient button for opt-out, I may increase the validity of Bayesian markets by mitigating bubbles in the market.

Even though Bayesian markets show promises in inducing truth-telling as a best response under the perfect belief, there is no benchmark for its validity in our experiment. Considering the requirement of sophisticated Bayesian reasoning in Bayesian markets, respondents may not respond to financial incentives by telling the truth. Even if they do, the improvement in truth-telling may not be enough to justify the application of Bayesian markets in practice. Therefore, I may better evaluate the validity of Bayesian markets with the help of a benchmark treatment, where subjects receive fixed payment in each period.

How well Bayesian markets perform relative to other truth-telling mechanisms is another related question. A natural candidate for comparison is BTS. In the same setup of priors and private information, it is possible to design comparable experiments of BTS. Instead of buying/selling assets, subjects are incentivized by information scores in BTS. Another possible candidate is the peer prediction method, which is possible due to the design of states and priors in our experiment, where mechanism implementers can learn common priors, even though it is not essential for Bayesian markets.

An important direction for future work is to test Bayesian markets with subjective truth. The experimental design is very challenging. On the one hand, tasks should be subjective enough for agents to believe that experimenters are impossible to verify the underlying truth. On the other hand, they should not be too subjective to be evaluated. One potential approach is to engage participants in tasks with partially subjective truth. For instance, it contains verifiable characteristics such as the range, mean, or distribution.

References

Arrow, K. J., Forsythe, R., Gorham, M., Hahn, R., Hanson, R., Ledyard, J. O., Levmore, S., Litan, R., Milgrom, P., Nelson, F. D., et al. (2008). The promise of prediction markets. *Science*, 320:877–878.

Baillon, A. (2016). A market to read minds. *Working Paper*.

Baillon, A. (2017). Bayesian markets to elicit private information. *Proceedings of the National Academy of Sciences*, 114(30):7958–7962.

Becker, G. M., DeGroot, M. H., and Marschak, J. (1964). Measuring utility by a single-response sequential method. *Behavioral Science*, 9(3):226–232.

Benjamin, D. J. (2019). Errors in probabilistic reasoning and judgment biases. In *Handbook of Behavioral Economics: Applications and Foundations 1*, volume 2, pages 69–186. Elsevier.

Berg, J., Forsythe, R., Nelson, F., and Rietz, T. (2008a). Results from a dozen years of election futures markets research. *Handbook of Experimental Economics Results*, 1:742–751.

Berg, J. E., Nelson, F. D., and Rietz, T. A. (2008b). Prediction market accuracy in the long run. *International Journal of Forecasting*, 24(2):285–300.

Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly weather review*, 78(1):1–3.

Camerer, C. F., Dreber, A., Forsell, E., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Almenberg, J., Altmejd, A., Chan, T., et al. (2016). Evaluating replicability of laboratory experiments in economics. *Science*, 351(6280):1433–1436.

Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Nave, G., Nosek, B. A., Pfeiffer, T., et al. (2018). Evaluating the replicability of social science experiments in nature and science between 2010 and 2015. *Nature Human Behaviour*, 2(9):637–644.

Dreber, A., Pfeiffer, T., Almenberg, J., Isaksson, S., Wilson, B., Chen, Y., Nosek, B. A., and Johannesson, M. (2015). Using prediction markets to estimate the reproducibility of scientific research. *Proceedings of the National Academy of Sciences*, 112(50):15343–15347.

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, 25(2):383–417.

Gao, X. A., Mao, A., Chen, Y., and Adams, R. P. (2014). Trick or treat: putting peer prediction to the test. In *Proceedings of the 15th ACM Conference on Economics and Computation*, pages 507–524. ACM.

Gneiting, T. and Katzfuss, M. (2014). Probabilistic forecasting. *Annual Review of Statistics and Its Application*, 1:125–151.

Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American statistical Association*, 102(477):359–378.

Good, I. J. (1950). Probability and the weighing of evidence.

Grether, D. M. (1980). Bayes rule as a descriptive model: The representativeness heuristic. *The Quarterly Journal of Economics*, 95(3):537–557.

Hayek, F. A. (1945). The use of knowledge in society. *American Economic Review*, 35(4):519–530.

John, L. K., Loewenstein, G., and Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, pages 524–532.

Jurca, R. and Faltings, B. (2006). Minimum payments that reward honest reputation feedback. In *Proceedings of the 7th ACM Conference on Electronic Commerce*, pages 190–199. ACM.

Jurca, R. and Faltings, B. (2009). Mechanisms for making crowds truthful. *Journal of Artificial Intelligence Research*, 34(1):209–253.

Manski, C. F. (2006). Interpreting the predictions of prediction markets. *Economics Letters*, 91(3):425–429.

Miller, N., Resnick, P., and Zeckhauser, R. (2005). Eliciting informative feedback: The peer-prediction method. *Management Science*, 51(9):1359–1373.

Offerman, T., Sonnemans, J., Van de Kuilen, G., and Wakker, P. P. (2009). A truth serum for non-bayesians: Correcting proper scoring rules for risk attitudes. *The Review of Economic Studies*, 76(4):1461–1489.

Parkes, D. C. and Witkowski, J. (2012). A robust bayesian truth serum for small populations. In *Proceedings of the 26th AAAI Conference on Artificial Intelligence*, pages 1492–1498.

Prelec, D. (2004). A bayesian truth serum for subjective data. *Science*, 306(5695):462–466.

Radanovic, G. and Faltings, B. (2013). A robust bayesian truth serum for non-binary signals. In *Proceedings of the 27th AAAI Conference on Artificial Intelligence*, pages 833–839.

Radanovic, G. and Faltings, B. (2014). Incentives for truthful information elicitation of continuous signals. In *Proceedings of the 28th AAAI Conference on Artificial Intelligence*, pages 770–776.

Savage, L. J. (1971). Elicitation of personal probabilities and expectations. *Journal of the American Statistical Association*, 66(336):783–801.

Schlag, K. H., Tremewan, J., and Van der Weele, J. J. (2015). A penny for your thoughts: a survey of methods for eliciting beliefs. *Experimental Economics*, 18(3):457–490.

Schotter, A. and Trevino, I. (2014). Belief elicitation in the laboratory. *Annual Review of Economics*, 6(1):103–128.

Shaw, A. D., Horton, J. J., and Chen, D. L. (2011). Designing incentives for inexpert human raters. In *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work*, pages 275–284. ACM.

Tziralis, G. and Tatsiopoulos, I. (2007). Prediction markets: An extended literature review. *The Journal of Prediction Markets*, 1(1):75–91.

Weaver, R. and Prelec, D. (2013). Creating truth-telling incentives with the bayesian truth serum. *Journal of Marketing Research*, 50(3):289–302.

Winkler, R. L. (1969). Scoring rules and the evaluation of probability assessors. *Journal of the American Statistical Association*, 64(327):1073–1078.

Winkler, R. L. and Murphy, A. H. (1968). good probability assessors. *Journal of applied Meteorology*, 7(5):751–758.

Wolfers, J. and Zitzewitz, E. (2006). Interpreting prediction market prices as probabilities. Technical report, National Bureau of Economic Research.