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

Yan Xu

First Version: October 2016
Latest Version

Abstract

Bayesian Market is a new mechanism that incentivizes individuals to tell the subjective truth. This paper tests the performance of Bayesian markets with three different degrees of manipulation in participants' beliefs in others' truthfulness. I find that Bayesian markets effectively induce truthful revelations when participants believe that others are truthful. However, when there is noise in agents' beliefs, Bayesian markets become less effective. Participants expect that the asset value is higher than the fundamental value and thus are more likely to buy assets, which further raising asset value ex-post. Since ex-ante belief of asset value is confirmed by its realization, bubbles arise in the market.

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

Thanks to....

*Erasmus School of Economics, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands. Email: xu@ese.eur.nl.

1 Introduction

Information elicited from dispersed individuals is increasingly crucial to many knowledge-gathering and decision-making tasks. For example, researchers are conducting extensive social-economic surveys to collect knowledge about human perceptions; crowdsourcing platforms are keen to solicit informative answers from online communities; more customers are steering their purchases to products with reliable customer reviews. In most cases, respondents are compensated for time and efforts, rather than for truthfulness; consequently, they may provide uninformative or even untruthful answers. A worker who labels images on MTurk[1], for instance, may type random tags. Alternatively, a respondent in a survey may lie about questions involving drug abuse or impaired driving. Hence, one of the biggest challenges of information elicitation is to design mechanisms that incentivize respondents to tell the truth.

When truth is available, many mechanisms create incentives for truth-telling. A popular one is Scoring Rule (Winkler (1969)), which is often employed to elicit individuals' probabilistic forecasts of a set of events. For instance, suppose we are interested in the event of whether there is rain or not on the next day. Survey-takers are first asked to submit probabilistic reports of rain. Then each report is graded against the objective truth – the actual proportion of times that rain fell – by a score. Typically, the closer a report to the underlying truth is, the higher the score it will attain and correspondingly the higher the reward will be. For more literature on comparisons and empirical implementations of various scoring rules, one can check recent reviews from Schlag et al. (2015) and Schotter and Trevino (2014).

Another popular mechanism with objective truth is Prediction Market (Arrow et al. (2008), Manski (2006)), wherein participants can buy or sell a contract paying $1 if a certain event occurs in the near future. For instance, in the 2016 US presidential election market[2], a share of "Buy Yes" contract associated with the question "Will Hillary Clinton win the 2016 U.S. presidential election?" will be redeemed at $1 if the prediction is true and 0 otherwise. Wolfers and Zitzewitz (2006) showed that resulting asset prices on prediction

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

[2] https://www.predictit.org/

markets are close to mean belief of all participants. Hence, it is optimal for agents who predict higher-than-average probabilities of the event to buy contracts and for those who predict lower-than-average to sell contracts.

Under both mechanisms, objective truth plays an important role in incentivizing truth-telling. In the example of scoring rule, truth of raining probability serves as an evaluation gauge for all possible reports; similarly, in the presidential prediction market, truth of election outcome is used to verify profits for all contracts. Thanks to underlying truth, payment schemes can be designed to yield the highest expected payoff for truthful reports and thus informants have incentives to tell the truth. Principally, validity of a mechanism hinges upon how well incentives of information providers are aligned to truthful revelations. When truth is available, such alignment can be easily achieved.

When truth is not available, it is less intuitive to apply the alignment principle. For questions like "Have you ever engaged in questionable research practice?", "Do you believe computers will outsmart humans?" or "Are you happy?", underlying truth is either subjective or costly to verify. Consequently, there is no natural benchmark against which respondents' answers are evaluated and truthful ones are rewarded. Nevertheless, subjective truth like feelings, judgements and emotions is more and more prominent in modern life. In this paper, we are primary interested in truth-telling mechanisms for which the ground truth is not accessible.

Recent research of mechanisms with subjective truth are categorized as the class of Peer Prediction and the class of Bayesian Truth Serum (BTS). The intuition is similar: incentive alignment can be achieved by constructing objective and verifiable benchmarks, which act as proxies for subjective truth. Different benchmarks, on the other hand, bring about differences in performance and applicability for these two classes of mechanisms.

Under Peer Prediction (Miller et al. (2005)), benchmarks for incentive alignment are peers' reports. Implementer of the mechanism, called Center, is assumed to be exposed to a common prior of private information and thus can calculate implied posterior belief about another participant (called reference)'s report. In the example of a survey question "Are you happy?", Center, by assumption, knows distribution of happy people in population and based upon which, she also knows the correct expectations that players should have about how many people are happy. Then each answer submitted by a particular player will be transformed into belief about a reference's report and further will be evaluated by a proper scoring rule. Since a reference's report is verifiable, truth-telling proves to be a Bayesian Nash Equilibrium. Insight of this scheme is quite clever: when common prior is known to Center, agents' private information can be verified by peer's information. Therefore, peers' reports act as reasonable surrogates for inherent truth and simple scoring rule will elicit truthful-telling. Jurca and Faltings (2006, 2009) further extended peer prediction method to prevent agents from coordinating on uninformative equilibria. However, they still required the assumption of omniscient Center, which is not likely to be satisfied in many practical scenarios.

Very few empirical or experimental work has been done to test peer prediction mechanisms. Gao et al. (2014) adopted the simplest setting of Jurca and Faltings (2009) in an experiment on MTurk, where participants repeated engaged in a task of reporting one of two signals based on different payment matrices. They found players were more likely to coordinate at uninformative equilibria, suggesting a failure of peer prediction in inducing the truth-telling equilibrium.

Bayesian Truth Serum (Prelec (2004)) relaxes the assumption of knowing common prior for Center. Under BTS, benchmarks are also constructed via peers' reports. However, each participant now needs to submit two reports, a signal report reflecting private information itself and a belief report corresponding to a prediction of the distribution of each answer. In the same example of "Are you happy?", each participant, in addition to answering "yes" or "no" to this question, has to predict the proportion of participants in large population answering "yes". Center, without necessarily knowing the common prior, assigns each participant a prediction score, which depends on how well the predicted distribution agrees with realized one, and an information score, which depends on how surprisingly common the submitted signal is in the realized signal distribution, as compared to the average prediction. The final payoff for each player is the weighted average of two scores. Since belief reports are verifiable, the prediction score induces truthful prediction of the distribution of signal report. The information score, on the other hand, exploits the implied Bayesian reasoning about population frequencies and rewards truthful signal reports that are more common than collectively predicted. Insight of this mechanism is similar to that of peer prediction: by exploring the relationship of private information among population, surrogates for subjective truth can be constructed to encourage truth-telling.

When population is large, truth-telling of both signal and belief reports is a Bayesian Nash Equilibrium. Parkes and Witkowski (2012) proposed Robust BTS for small population. Furthermore, Radanovic and Faltings (2013, 2014) generalized RBTS to non-binary and continuous signals.

Empirical test of BTS is also scant. John et al. (2012) conducted an anonymous survey of 2000 psychologists about their involvement in questionable research practices, using BTS as incentive schemes. They found that BTS induced 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. Despite these encouraging evidences, the validity of BTS in practical implementation remains an open question. In above two experiments, subjects were neither taught formulas of scores nor the notion of equilibrium. Instead, they were suggested to believe that truth-telling was in their best interest. Since subjects could not easily link their actions to payoffs, their incentives of truth-telling might be distorted. For instance, Shaw et al. (2011) employed BTS as a contextual manipulation on MTurk and they found that workers performed significantly better, even though they were not financially rewarded by BTS. They argued that outperformance of BTS might be attributed to confusion and cognitive demand.

Baillon (2017) proposed a new institution, called Bayesian Market, to simplify practical implementation of BTS. Through a market where private information is linked to asset transactions, individuals are rewarded for 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 and a corresponding bid (ask) price. Then asset value and asset price are determined. Asset value is the realized proportion of people answering "Yes" to this question and asset price is randomly determined by BDM method (Becker et al. (1964)). Similar to information scores 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. Moreover, this equilibrium is stable against disturbances of belief, meaning even if agents expect some other agents to lie, truth-telling remains optimal.

This paper aims to test 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 demand discussed in Shaw et al. (2011). Another one inherits in the definition of truth-telling BNE, where reporting truthfully is optimal for an agent if he believes all other agents are truthful. Even if allowing for some disturbances, this belief is not easy to be satisfied in real scenarios. To understand how effective Bayesian markets will be under different belief systems, we distort individuals' belief and gradually relax its perfectness. To be specific, we plan to answer 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 even when participants are allowed to expect some agents to lie? And (3) will Bayesian markets induce the truth-telling BNE?

We achieve these goals by constructing three types of laboratory Bayesian markets, each featuring a setting where participants' belief of others' truthfulness is distorted to some degrees. Inspired by Gao et al. (2014), we create links of private information among agents through bingo cages, which imply common priors and further generate private signals for all agents. After a draw, an agent updates his belief in the probability of each type of signals in population and determines his trade position and the corresponding price in a Bayesian market. Control over belief, on the other hand, is achieved by virtue of Algorithm Agent (AA)s, who always report their own private signals and their correct belief in asset value. By varying proportions of AAs among eight agents in Bayesian markets, we expose human agent (HA)s to different degrees of belief in truthfulness. Particularly, there are 3 treatments – 1HA, 3HA and 8HA treatment, each aiming to address one corresponding research question proposed before. Therefore, Bayesian markets can be tested extensively for their validities in inducing truth-telling both as a best response and as a BNE.

We found Bayesian markets effectively induced truthful reports of private signals and posterior expectations of asset value under perfect belief systems. But when there are disturbances of agents' belief in others' truthfulness, Bayesian markets were less effective. These treatment effects can be explained by the arise of bubbles in the market. Due to noise in belief, agents who receive blue ball signals are more likely to buy than to short sell assets, which causes expectations of asset value higher than 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 two, we briefly describe Bayesian market mechanism and its theoretical predictions. In section three, the experimental design and procedures are introduced. Section four is data analysis of experiments. In section five, we conclude this paper and discuss potential future works.

2 Bayesian Market Mechanism

2.1 Model Setting

We begin with a formal game theoretical framework for elicitation mechanisms. To be consistent with the experimental design, we consider the simplest case that n homogeneous risk neutral agents are surveyed with the same binary question as before – "Are you happy?". Population n is further assumed to be infinite for theoretical simplicity in this section. But it will be relaxed in the following section. Private information of agent i is a random variable, denoted as T_i. Once realized, agent i learns his signal (or type) t_i, which takes value from $\mathcal{T} = \{Y, N\}$. The most important object for a truth-telling mechanism is to induce agents to truthfully report their signals.

Before receiving private signals, agent i believes that signals are drawn from a joint distribution $f(T_1, T_2, ..., T_n)$. We assume that this prior distribution is same for all agents and is common knowledge. In the case of binary signals, a common prior can also be described as $f(\omega)$, where ω is the proportion of agents whose private signal is Y. Given a realization of T_i as t_i, agent i updates his belief in ω through $f(\omega|T_i = t_i)$. Under a mechanism like peer prediction, every agent is required to submit a report and Center needs to specify payoffs for all possible report profiles. Thus, a mechanism can be defined by a pair $(\mathcal{R}, \mathcal{M})$, where $\mathcal{R}$ is the set of all possible reports from individuals and $\mathcal{M} : \mathcal{R}^n \to \mathbb{R}^n$ is a mapping from a space of report profiles to that of payoff profiles. When a mechanism requires agents to submit several reports, like in the case of BTS, $\mathcal{R}$ will be high dimensional.

The strategy of agent i under mechanism $\mathcal{M}$ is defined as a mapping $s_i : \mathcal{T} \to \Delta\mathcal{R}$, where $\Delta\mathcal{R}$ is a probability distribution over $\mathcal{R}$. A strategy profile s is the collection of strategies of all agents $(s_1, s_2, ..., s_n)$ and $\mathcal{S}$ is the set of all such profiles. The utility of agent i is defined as $u_i(s_i(t_i), s_{-i}(t_{-i}))$, where $s_{-i}(\cdot)$ is a strategy vector of all agents but i.

A strategy profile $(s_1^*, s_2^*, ..., s_n^*)$ is a BNE under mechanism $\mathcal{M}$ if:

$$E_{T_{-i}}[u_i(s_i^*(t_i), s_{-i}^*(T_{-i}))|T_i = t_i] \geq E_{T_{-i}}[u_i(s_i(t_i), s_{-i}^*(T_{-i}))|T_i = t_i]$$

for all $i \in \{1, 2, ..., n\}$ and all $s_i \in \Delta\mathcal{R}_i$. With inequality, each agent maximizes his expected utility by following strategy s_i^* given his own signal and given that the other agents play according to strategies s_{-i}^* in the equilibrium.

2.2 How Bayesian Markets Work?

In a Bayesian market, each agent has an opportunity to trade at most one asset by submitting a yes/no report and a corresponding bid/ask price. First of all, signal reports are inferred from individuals' trading positions on Bayesian markets. If an agent decides to be a buyer, a yes report ($r_i = Y$) will be automatically sent to a market maker. Similarly, being a seller is equivalent to submitting a no report ($r_i = N$). In addition, belief reports are elicited through bid/ask prices: 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 a task of asset transaction, which is more familiar for them and thus may create greater engagement.

To be consistent with the general model setting, we define report space for agent i as $\mathcal{R}_i = \{(r_i, c_i)\}$, where $r_i \in \{Y, N\}$ is a signal report and $c_i = \mathbb{1}_{\{r_i=Y\}}(b_i - a_i) + a_i$ is a belief report. A report profile for all agents can be represented by $(r, c) = \{(r_i, c_i)\}_{i=1}^n \in \mathcal{R}^n$.

Value associates with the asset, denoted as v, is the proportion of yes reports among those who participate in the market. Under full participation[3], $v = \frac{1}{n}\sum_{j=1}^n \mathbb{1}_{\{r_j=Y\}}$. The market price of the asset, denoted as p, is randomly drawn from a commonly known distribution $g(p)$ on $[0, 1]$.To prevent agents from learning private information through direct trade, there is a market maker between buyers and sellers. She collects all reports from participants and calculates the average bid and ask price as her buying ($\bar{a}$) and selling ($\bar{b}$) price. The trading mechanism between agents and the market maker follows BDM method. Specifically, a trade occurs for buyer i if $b_i \geq p \geq \bar{a}$, under which both

[3] In the truth-telling equilibrium, it is indeed the case.

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 0. Mathematically, Bayesian markets are thus characterized by the following payoff function:

$$\mathcal{M}_i(r, c) = \frac{1}{n}(\mathbb{1}_{\{r_j=Y\}}\mathbb{1}_{\{b_i \geq p \geq \bar{a}\}}(v - p) + \mathbb{1}_{\{r_j=N\}}\mathbb{1}_{\{a_i \leq p \leq \bar{b}\}}(p - v))$$

2.3 Properties of Bayesian Markets

A Bayesian market's validity in truth-inducing relies on the link between an agent's private signal and his expectation of others' signals. Hence, we start theoretical analyses by investigating how agents formulate and update belief in ω in Bayesian markets.

Before receiving private signals, agents share a common prior $f(\omega)$ about the proportion of Y-type in population. This assumption is quite strong, but is pervasively assumed in truth-telling mechanisms. After receiving a private signal t_i, agent i updates his belief in the proportion of Y-type based on Bayes rule, denoted as $f(\omega|t_i)$. Signals are assumed to be "impersonally informative". By "informative", we expect signals to provide information about population frequency ω; by "impersonal", we admit that agents who receive same signals will learn in the same way of ω. Mathematically speaking, we have:

$$f(\omega|t_i) = f(\omega|t_j) \Leftrightarrow t_i = t_j,$$

which implies both "informative" and "impersonal" property. Condition $f(\omega|t_i) = f(\omega|t_j) \Rightarrow t_i = t_j$ means that different types of agents will formulate different posterior belief. On the other hand, condition $f(\omega|t_i) = f(\omega|t_j) \Leftarrow t_i = t_j$ requires that respondents of same type will draw same inference of ω.

Given signal and belief structures, we can further explore how agents make decisions and the resulting equilibrium in Bayesian markets. Since this paper is more focused on testing the mechanism, we will just summarize key results here. For formal proofs and other details, one can refer to Baillon [2016].

Proposition 1. *Y-type agents expect a higher frequency of Y-type signal in population than N-type agents. Namely,* $E(\omega|t_Y) > E(\omega|t_N)$.

This proposition characterizes Bayesian reasoning of belief updating and thus forms the basis for incentive alignment in Bayesian markets. When receiving different private information, agents formulate different expectations of others' private information. In particular, agents who are truly happy (Y-type) will expect more happy people in population than those who are not (N-type). It should be noted that it is a relative, rather than an absolute comparison. For example, both types of agents may expect a minority of N-type in population, however, by exploiting the information contained in their true answers, happy agents still expect that Y-type signals are more common in population than what unhappy agents expect.

Proposition 2. *Agent who participates in a Bayesian market will submit his posterior expectation of asset value* $E(v|t_i)$ *as his bid or ask price.*

This proposition means that contingent on agent i's buy/sell decision, the willingness to pay for an asset is $b_i = E(v|t_i)$ and the willingness to accept an asset is $a_i = E(v|t_i)$. After deciding trade positions on markets, buyers/sellers choose bid/ask prices to maximize posterior expected payoff and the optimal one happens to be $E(v|t_i)$. This result 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 also equals to the posterior expectation of asset value.

For notation simplicity, posterior expectations of different types of agents are represented by $E(v|t_Y) \equiv \omega_Y$ and $E(v|t_N) \equiv \omega_N$, respectively. Based on Proposition 2, we further calculate a market maker's buying and selling price if all agents follow the optimal bid and ask strategy. When the number of agents n goes to infinity, the market maker's buying price is $\bar{b} = \omega_Y$ and selling price is $\bar{a} = \omega_N$.

Proposition 3. *Truth-telling is a BNE in Bayesian markets.*

This proposition predicts the truth-inducing property of Bayesian markets. In fact, it is jointly implied by the definition of BNE, Proposition 1 and 2. First, by BNE, when all other agents truthfully report private signals, asset value on the market is the same as the frequency of Y-type signal (ω). Then, according to Proposition 2, active agents in the market will report their posterior expectations of asset value, producing buying price ω_Y and selling price ω_N for a market maker. From Proposition 1, a Y-type agent expects a higher asset value than a N-type agent and thus is willing to buy an asset. Similarly, a N-type agent is willing to sell an asset. In the equilibrium, both types of agents will participate in markets and will truthfully reveal their private signals by their buy/sell decisions.

Proposition 4. *Proposition 1-3 still hold for small samples with at least 4 agents.*

This proposition is particularly useful for practical implementation of Bayesian markets. One problem of small sample is the possible manipulation of asset value through one's own report, which may encourage deception. To ensure incentive alignment of both signal and prediction reports, assets on Bayesian markets are individually adjusted. Specifically, each agent chooses to buy/sell an "individualized" asset whose value is the proportion of yes report among all other agents on the market[4]. In addition, there will be no trade for an individual when all other agents submit same signal reports. With these adjustments, all above results still hold: BDM method incentivizes agents to submit truthful prediction reports; Y-type agents expect a higher asset value than N-type agents; truthful report of singles for every agent in an equilibrium.

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 of which are directly defined on the signal spaces. Unlike mechanisms like peer prediction, Bayesian markets do not require the structure of how private information is

[4]Asset value is defined in this way to be consistent with the experimental design in the next section. Theoretically, three randomly selected reports from other agents are enough to support Proposition 4.

formed. In practice, it yields convenience. For example, we may don't know 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 is more focuses on whether and how Bayesian markets induce truth-telling, we design the experiment in accordance with the simpler case where both common priors and signals are determined by states of the world.

There are two states of the world, which is 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, we 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 states 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 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:

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

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.

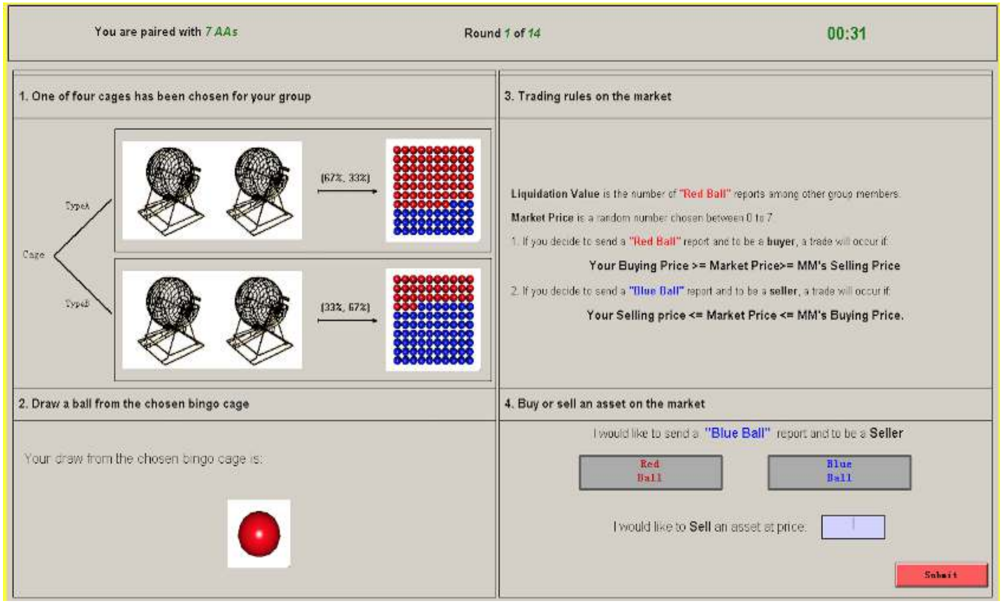


Figure 1: Decision Screen

Based on that, he further updates 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. We 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 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, we try to increase the chance of transactions between traders and market makers in the experiment.

We 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, we introduce learning as a possible de-bias device and focus on testing whether Bayesian markets will induce truth-telling in convergence. To facilitate learning, we keep the group members fixed during the whole session and provide each subject full information of signals, decisions and profits after current period completes. An example of review screens is depicted in Figure 2.

Figure 2: Review Screen

It is important to emphasis that Bayesian markets rewards 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 actually 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 hold such belief. For example, it's reasonable for an 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 to tell the truth. In order to understand whether and how Bayesian markets perfoms, we need to design the experiment in a way that subjects' belief can be controlled.

Our approach is to 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 we are interested in. AAs, on the other hand, are programmed to truthfully report private signals and posterior expectations of asset value. To make a participant believe that all other agents are truth-telling, we can group him with AAs. In other words, a HA's belief of other agents' truthfulness can be controlled by the number of HAs and AAs in the market. Following this simple idea, we design three treatments, which differ in the degree of others' truthfulness that a participant believes.

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, we 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 2 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]. We 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 7 other HAs. They may form different belief 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

We 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 is 21.80 euro. The number of subjects, groups and average payoffs for each treatment are described in Table 1:

Table 1: Treatments

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

Upon arrival, each subject were randomly assigned an ID and were 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 informa-

[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[2016]

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

tion. At the end of the experiment, we 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 token, each worth 0.5 euro. After a market had closed in one period, asset price was randomly chosen from (0,7). Then a market 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 loss money, they were endowed with 3 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, we 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 Data Analysis

The analyses of experimental data are mainly focused on three aspects: truthful rate, bid and ask prices and possible heuristics.

4.1 Truthful Rate

4.1.1 Aggregate Truthful Rate

We first turn to aggregate truthful rate, which is the proportion of truthful reports among all observations. Table 2 summarizes average percentages of truthful reports submitted by agents in three treatments.

Unsurprisingly, all three truthful rates are lower than the theoretical value, which is 100%. Even though our experiment satisfies prior and information assumptions of Bayesian mar-

Table 2: Aggregate Truthful Rate

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

kets, 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 be able 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 still reveal valuable information about the performance of Bayesian markets . In 1HA treatment, 80% of the reports submitted by agents were same as private signals. But this number was much lower in 3HA and 8HA treatment. Mann-Witney tests showed that the truthful rate in 1HA treatment was significantly higher than that in 3HA and 8HA treatment, but there was no significant difference between 3HA and 8HA. This implies that when there are more HAs in the market, participants may feel uncertain about whether others will be truthful or not, and thus Bayesian markets are less effective in inducing truth-telling.

4.1.2 Individual Truthful Rate

We also calculated individual truthful rate for each treatment, which is the frequency of truth-telling for each subject. The histogram of three individual truthful rates are depicted in Figure 3.

Similarly, there are significant treatment effects. In 1HA treatment, 25% of subjects reported their private signals truthfully and 21% only lied once during 14 periods. However, in 3HA and 8HA treatment, few subjects were fully truthful and as a consequence, aggregate truthful rates are lower than that in 1HA treatment. In terms of inducing truth-telling from individuals, again, we found Bayesian markets outperformed in 1HA treatment.

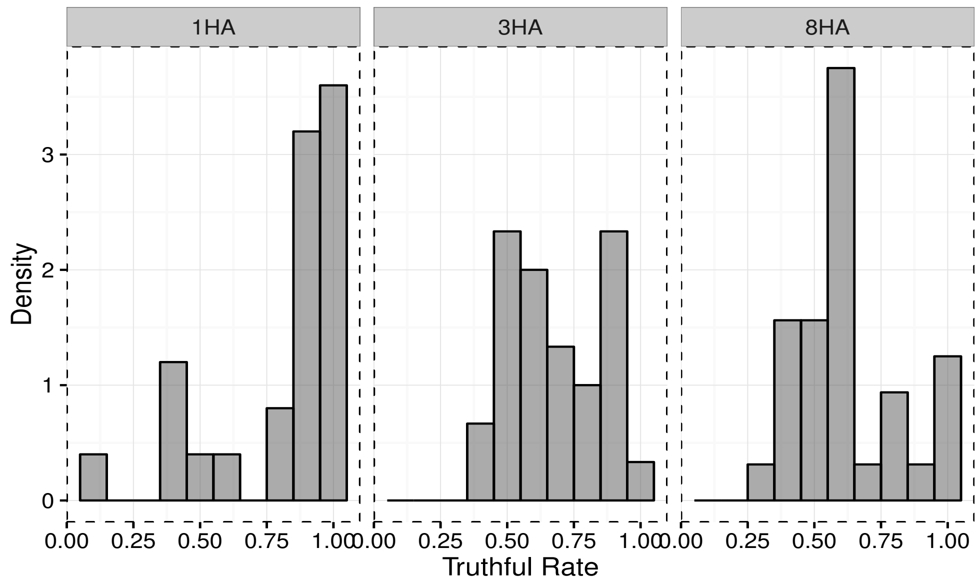


Figure 3: Histogram of Individual Truthful Rate

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

4.1.3 Heterogeneity in Strategy Evolutions

One possible source of treatment effects is the heterogeneity in strategy evolutions 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 was increasing with the number of HAs in the market, we expected learning speed to be the highest in 1HA treatment, medium in 3HA and the lowest in 8HA treatment.

Figure 4 illustrates time series of average truthful rate in three treatments. The solid line is for 1HA treatment, dotted for 3HA and dashed for 8HA treatment. Surprisingly, there was little evolution trend in both 1HA and 8HA treatment. For 1HA treatment, the truthful rate was persistently high, which might limit space for further improvement. For 8HA treatment, the interaction among human agents might introduce too much noise for effective learning. However, there was 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.

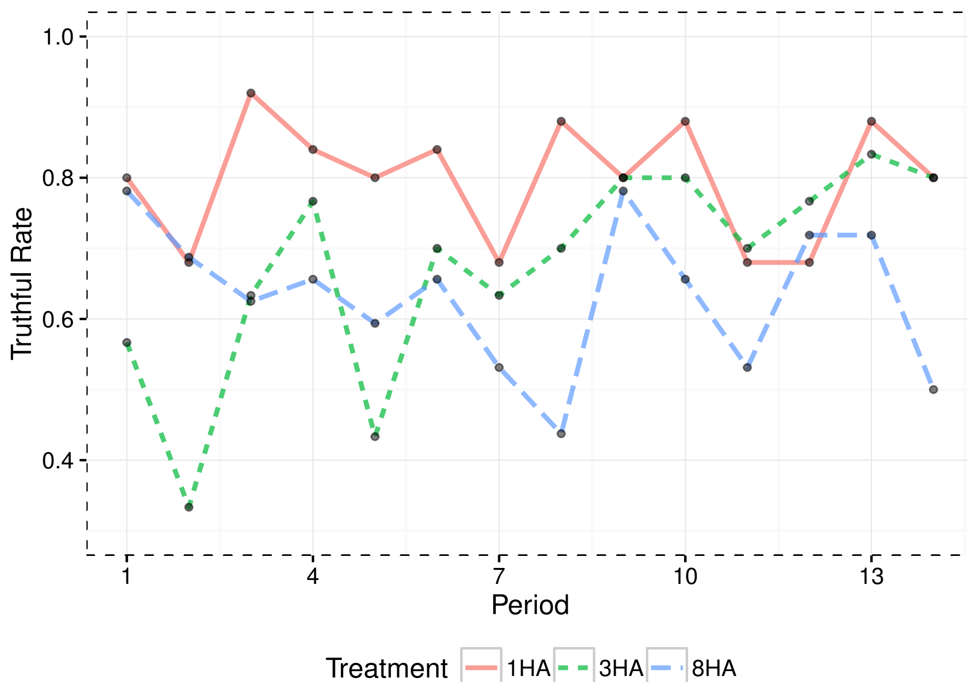


Figure 4: Truthful Rate by Periods

We did find heterogeneity in strategy evolutions among treatments, but it was 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 learning in 3HA. However, strategy evolution failed to explain treatment effects between 1HA and 8HA treatment, in which truthful rates both evolved steadily at different levels.

4.1.4 Heterogeneity in Signal Effects

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

Table 3: Posterior Truthful Rate

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

At first sight, this result was quite puzzling because different labels of private signals should not have affected truthful rates in a systematic way. Moreover, our experiment was designed to be symmetric of signals: first, it was equally likely to receive a red ball or a blue ball on average; second, each market in the experiment corresponded to another market where signal labels were switched. Therefore, posterior truthful rates should be equal for both types of signals.

However, inherit elicitation procedures of Bayesian markets might attach different perceptions to different signals. Specifically, signal reports submitted by agents were associated with their buy/sell decisions. Implied transaction positions of private signals, rather than labels, might affect incentives of truth-telling. Since these incentives were triggered by market conditions, they were highly likely to vary with group compositions for each treatments. A plausible explanation was that agents were more likely to buy than to short sell assets because they were more experienced with purchase decisions. This phenomena is quite normal in market institutions and we call it buying inclination here. In our experiment, as the number of HAs increased, markets became more noisy, making the choice of buying/selling more difficult and thus resulting in a more severe buying inclination.

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

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

Table 4: Buying Rate and Truthful Rate

	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

Since the aggregate truthful rate was an average of truthful rates of buyers and sellers with the buying rate as a weight, we concluded that treatment effects were driven by signal effects. More specifically, two factors jointly played a role: First, agents in 3HA and 8HA treatment were more likely to buy than to sell assets; second, buyers were more likely to lie than sellers.

4.2 Bid and Ask Prices

4.2.1 Aggregate Bid/Ask Prices

Figure 5 demonstrates time series of submitted price, theoretical price and asset value in Bayesian markets. The solid line is the average prices submitted by subjects in each period, representing an average HA's ex-ante prediction of asset value. The correct predictions based on private signals are shown by the dashed line. And the dotted line is the ex-post asset value.

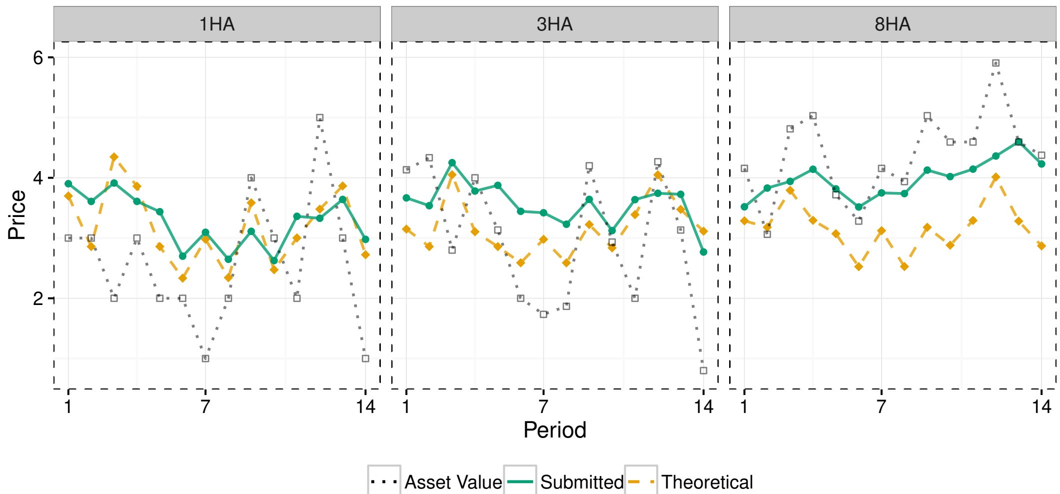


Figure 5: Aggregate Bid/Ask Price and Theoretical Price

We focused on both ex-ante and ex-post prediction gaps in each treatment. The first one means the difference between submitted and theoretical prices and the second one is the difference between submitted prices and realizations of asset value.

First, ex-ante prediction gaps in all treatments were positive for all 14 periods, meaning that an average HA in a Bayesian market predicted higher asset value than a AA did. Therefore, HAs were more likely to buy assets than AAs in all treatments, which was consistent with buying inclination shown in previous subsection.

In terms of size and trend of ex-ante prediction gaps, there was a significant difference between 1HA (3HA) and 8HA treatment. In both 1HA and 3HA treatments, ex-ante prediction gaps were quite small and showed no trend. In other words, an average HA in these two treatments almost correctly predicted asset value, ex ante. While in 8HA treatment, the prediction gap started at a substantial level and was increasing over time, indicating that HAs were persistently over-predicting asset value.

As for ex-post prediction gaps, there was also a significant difference between 1HA (3HA) and 8HA, but the direction was opposite: An average HA in 8HA treatment predicted more accurately of realizations of asset value than those in 1HA and 3HA treatment.

The treatment effects of both ex-ante and ex-post prediction gaps were consistent with each other and jointly explained the trend-chasing of submitted prices in 8HA treatment. Since subjects in 8HA treatment over-predicted asset value more severely, they were over-buying assets in the markets, which further drove up the ex-post asset value. In other words, their belief was self-confirmed by their decisions, leading to bubbles in Bayesian markets. In the following subsections, we to understand what caused bubbles in 8HA treatment.

4.2.2 Bid/Ask Prices for Truthful and Untruthful Reports

The first conjecture was that untruthful subjects might be the main drive for the over-prediction of asset value and for the resulting over-buying in markets. This was confirmed by Figure 6, where we depicted the same time series for both truthful and untruthful reports.

On average, truthful agents submitted prices close to theoretical prices. They acted as Bayesians who exploited private information to update belief. Untruthful agents, on the other hand, severely over-predicted asset value. They acted as trend-chasers who took into account possible bubbles in the market. For 8HA treatment, comparisons between truthful and untruthful agents were more evident: truthful subjects submitted predictions reflecting fundamental value of the asset and untruthful subjects chased bubbles and submitted predictions reflecting market valuation of the asset.

4.2.3 Price Divergences in 8HA Treatment

Since untruthful reports were in the form of either receiving a red ball but reporting blue ("RB") or receiving a blue ball but reporting red ("BR"). We further checked time series of submitted price, theoretical prices and asset value based on private signals and buy/sell decisions in 8HA treatment in Figure 7.

Figure 6: Average Bid/Ask Price and Theoretical Price by Truthfulness

The same pattern appeared: agents who received blue ball signals and who decided to be buyers are trend-chasers whose submitted prices are close to the realization of asset value; agents who received red ball signals and who decided to be sellers acted as Bayesians and submitted correct belief in asset value based on private information.

Combining with the buying inclination in previous section, we found that agents in 8HA treatment who received blue ball signals predicted higher asset value than fundamental

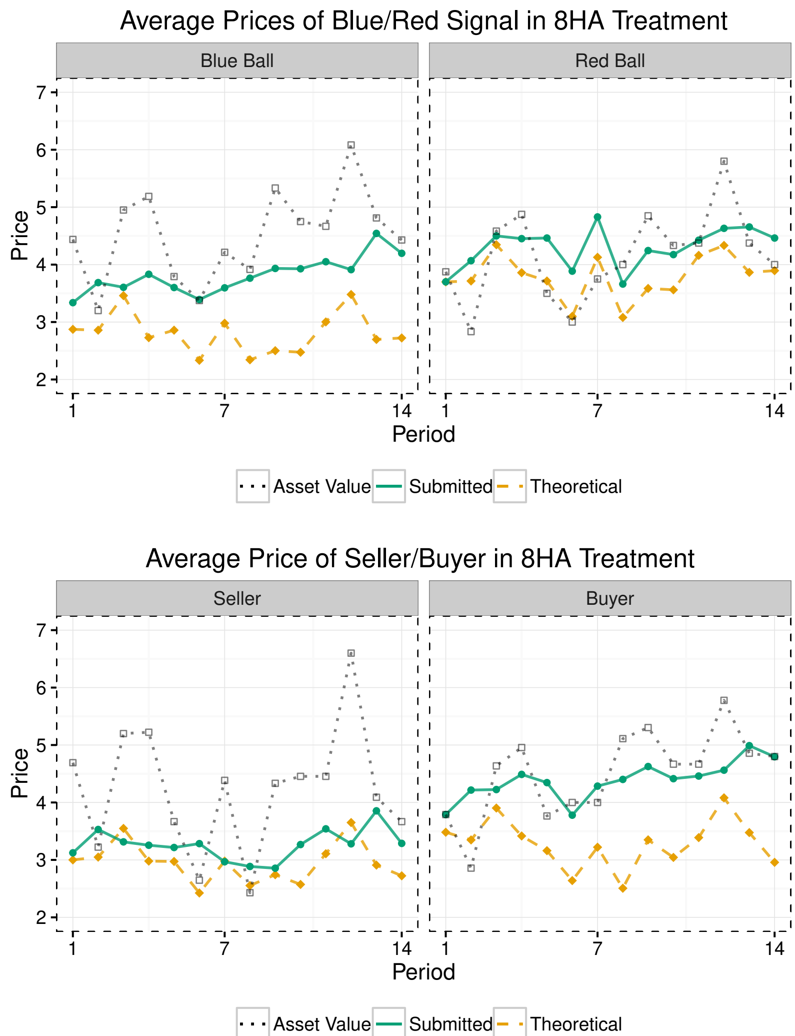


Figure 7: Price Divergence in 8HA Treatment by Signals and Buyer/Seller Decisions

value and thus were more likely to buy assets, which further raised the realized asset value. Since their belief in asset value was confirmed by its realization, they would continue to chase the trend and finally resulted in bubbles in the market.

Another point to notice is that all time series were the aggregate of four Bayesian markets, each corresponding to a group of 8HAs in the treatment. Appendix III showed same time series of submitted price, theoretical price and asset value for each group. We found that aggregate ex-ante prediction gaps were mainly driven by Group 24. In all other 3 markets, there was no clear evidence of trend-chasing and the resulting market bubbles. If Group 24 was deleted, the extend of bubbles in 8HA treatment would be largely reduced.

4.3 Heuristics

4.3.1 Keep and Switch Heuristics

A simple heuristic of learning is keeping or switching strategies conditional on profits in previous period. Under this heuristic, subjects may follow the strategy of previous period if it yielded positive profit and switched to alternative strategies otherwise.

Figure 8 captures frequencies of switch and keep patterns for each treatment. There were 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. We found that truth-telling was quite focal in all treatments. The frequency of "Keep truth-telling" stayed steadily around 70% in 1HA treatment but was more volatile in 3HA and 8HA treatment. Remarkably, there was an increasing trend in 3HA for the pattern of "Keep truth-telling", which explained the learning effect of truthful rate in 3HA treatment. All other three pattens were much volatile.

The change of frequencies of switch patterns might relied on previous profits. We listed the conditional switch rate for each treatment inTable 5. We found more switches in both 3HA treatment and 8HA treatment. However, conditional on previous profits, there was no significant difference between switch frequencies. Collectively, when there was noise of belief in others' truthfulness, subjects did try alternative strategies, switching from truth-telling to lying or from lying to truth-telling. However, these switches did not depend on previous profit.

4.3.2 Imitation Heuristic

Another simple heuristic is to imitate strategies that yielded the highest profit in previous round. In our experiment, when each period completed, subjects could learn all information

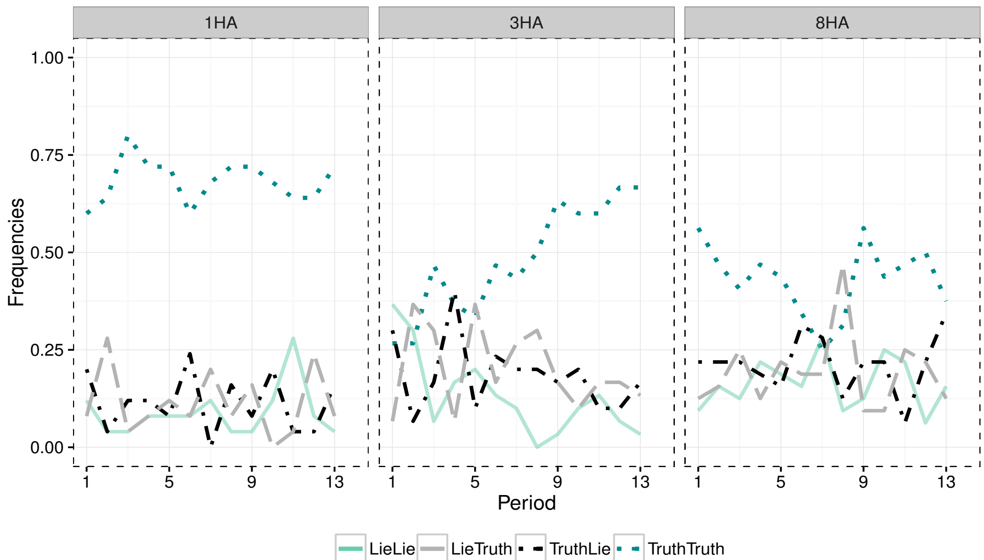


Figure 8: Time Series of Strategy Keep and Switch

Table 5: Switch Heuristic

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

of the previous period on a review screen. They may check strategies and the corresponding profits of others and further adjust his strategy in next round. Table 6 reports the imitation rate and winner's truthful rate for each treatment. Imitation rate was calculated as the frequency of which a subject's current strategy was the same as the winner's strategy in previous round.

First, winner's truthful rate winner's truthful rates were close to 1 in 1HA and 3HA treatment, implying truth-telling yielded the highest profit on average. This was not surprising because of the existence of AAs in these two treatments, who could also be winners, were al-

Table 6: Imitation Rate

Treatment	1HA	3HA	8HA
Imitation Rate	0.80	0.67	0.55
Winner's Truthful Rate	0.99	0.96	0.69

ways truthful. However, in 8HA treatment, without AAs to stabilize market price, bubbles might arise and lying could be the winning strategy. Hence the truthful rate for winners was significantly lower.

Second, imitation rate in 1HA treatment was 80%, significantly higher than that in 3HA and 8HA treatment. It should be noted that imitation rate here captured coincidence, rather than causality between subjects' strategy and winners' strategy. Since it was impossible to isolate imitation from other heuristic or strategic considerations solely from revealed decisions, we could not conclude that subjects in 1HA were more likely to imitate winners' strategy. But imitate rates still showed a higher correlation between subjects' strategy and winner's strategy in 1HA treatment.

To study how imitation heuristic affected truth-telling in each treatment, we depicted the scatter plot of imitation rate and truthful rate in Figure 9. The correlation was quite strong in 1HA and 3HA treatment. But in 8HA treatment, high imitation rate might not imply (or be implied by) high truthful rate because the winning strategy was less likely to be truthful.

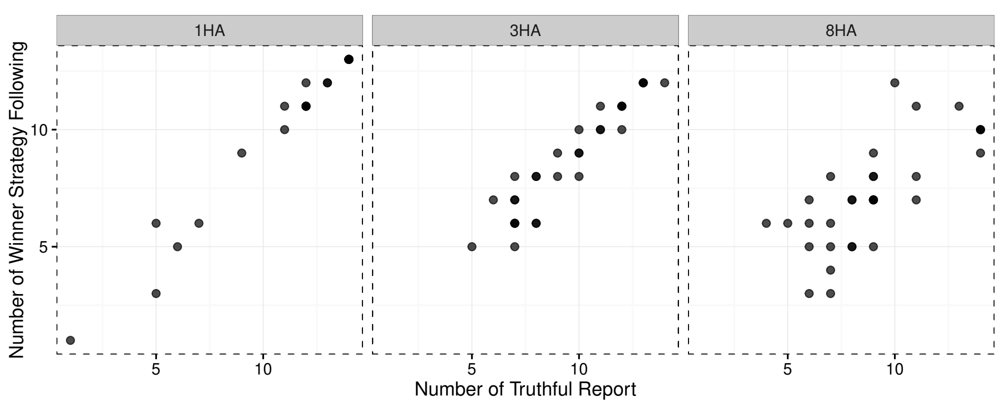


Figure 9: Relationship between Imitation Rate and truthful Rate

5 Conclusion and Discussion

This paper aimed to test whether Bayesian markets would induce truth-telling, either as a best response or as an equilibrium, from individuals. We achieved this by an experiment where Bayesian markets were constructed and human agents' belief in others players' truthfulness was manipulated in three settings.

First, we found that Bayesian markets effectively induced truth-telling as a best response when agents were reasonably to believe that all their opponents were truthful. In particular, agents submitted truthful reports of private signals in most cases (80%) and formed correct posterior expectations of asset value. However, when agents believed that some other agents might lie, Bayesian markets were less effective. Participants in the 3HA and 8HA treatment were less likely to reveal private information and they over-predicted asset value.

Second, we found that bubbles might arise in Bayesian markets, especially when there were more noise to agents' belief in others' truthfulness. We interpreted this phenomena as follows: (1) Common priors and impersonal private information were not sufficient to induce common posterior expectations; (2) traders who were supposed to sell assets predicted higher - than - fundamental asset value and thus were more likely to buy assets; (3) over-buying in the market raised realization value of assets, which further confirmed ex-ante belief; (4) these traders would further chase trend to buy assets and thus bubbles were made possible.

A closer check of what generated the bubble in the first place revealed buying inclination among participants, which also explained relatively lower frequency of truth-telling in 3HA and 8HA treatment. To be specific, when traders were more uncertain about behaviors of other traders, they would be more likely to buy than to short sell assets, simply because of the familiarity of the former context. Initially, it might just raise average expectation of asset value slightly, but it had the potential to trigger self-confirming belief in the market and to bring about bubbles. One important lesson for the practical implementation of Bayesian market mechanism is that we 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 because truth-telling yields strictly positive payoff in expectation. However, when an agent believes that the market is out off 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 loss. 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). They might not realize there was an option to step out of the market. By introducing treatment with a salient button for opt-out, we may increase the validity of Bayesian markets by mitigating bubbles in the market.

Even though Bayesian markets showed promises in inducing truth-telling as a best response under perfect belief, there was 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 truth. Even if they do, the improvement in truth-telling may not be enough to justify the application of Bayesian markets in practice. Therefore, we may better evaluate the validity of Bayesian markets with the help of a benchmark treatment, where subjects receive fixed payment in each period.

A related question is how well Bayesian markets perform relative to other truth-telling mechanisms. A natural candidate for comparison is BTS. In the same setting of priors and private information, we can 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. With the same setting of states and priors, we can design experiments for peer prediction.

We believe that an important direction for future work is to test Bayesian markets with subjective truth. The experimental design is very challenging. On 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 including its 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. (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.

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.

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.

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.

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.

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.

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

Appendix I: Parameters

CA is the probability that the chosen cage for the group is of Type A.
RA (RB) is the proportion of red ball in a cage of Type A (B).
w is the prior of the proportion of red ball signals among population.
wR (wB) is the posterior of the red ball proportion among population conditional on receiving a Red (Blue) Ball signal.
AAbid (AA ask) is the bid (ask) price submitted by AAs, which is adjusted from wR (wB) due to small sample restriction in the experiment.

Table 7: Parameters for Priors

	CA	RA	RB	w	wR	wB	AAbid	AAask
1	0.50	0.70	0.30	0.50	0.58	0.42	0.55	0.39
2	0.25	0.85	0.38	0.50	0.58	0.42	0.55	0.39
3	0.75	0.62	0.15	0.50	0.58	0.42	0.56	0.39
4	0.50	0.67	0.33	0.50	0.56	0.44	0.53	0.41
5	0.25	0.80	0.40	0.50	0.56	0.44	0.53	0.41
6	0.75	0.60	0.20	0.50	0.56	0.44	0.53	0.41
7	0.50	0.57	0.23	0.40	0.47	0.35	0.44	0.33
8	0.25	0.70	0.30	0.40	0.48	0.35	0.44	0.33
9	0.50	0.77	0.43	0.60	0.65	0.53	0.62	0.50
10	0.75	0.70	0.30	0.60	0.65	0.52	0.62	0.49
11	0.50	0.65	0.25	0.45	0.54	0.38	0.51	0.35
12	0.25	0.80	0.34	0.46	0.54	0.38	0.51	0.36
13	0.50	0.75	0.35	0.55	0.62	0.46	0.59	0.43
14	0.75	0.66	0.20	0.55	0.62	0.46	0.59	0.43

Appendix II: Experimental Instruction (1HA)

Preliminary Remarks

Your payoff in today's experiment is directly linked to your decisions. Please pay careful attention to this instruction because it can help you to better understand the experiment.

During the experiment, please do not communicate with any other participants and do not look at their computer monitors. If you have questions at any time, please raise your hand and we will come to you promptly.

Overview

The currency in the experiment is token and the conversion rate between tokens and euros is: 1 euro = 4 tokens. The experiment consists of 20 rounds. In each round, you will be endowed with 4 tokens to participate in a market to trade an asset. Your tasks are to: (1) decide whether to buy or sell an asset; (2) provide prices at which you would like to buy or sell the asset. These two decisions jointly influence the possibility of whether a trade will occur and the associated profits. Your payoff in each round is the sum of the endowment and your profit. The total payoff in the experiment is the sum of payoffs for the 20 rounds.

During the experiment, you will be grouped with seven Algorithm Agents (AAs). They are programmed identically and will face the same tasks as you do.

There is a "Market Maker" (MM) on the market, who collects all agents' decisions in your group and formulates his buying/selling price. You can trade at most one asset with MM.

At the end of each round, the asset will be liquidated[8]. If there is no trade, your profit is 0. If a trade occurs, a buyer's profit is the difference between the *Liquidation Value* and the *Market Price* and a seller's profit is the difference between the *Market Price* and the *Liquidation Value* .

At the end of the experiment, you will be paid privately. When we call your ID number, please come forward to the sign-in counter and receive your earnings. We ensure you that any information regarding your participation, your name, ID number and payoffs are kept strictly confidential.

How does the Market Work?

What is the Structure of the Market?

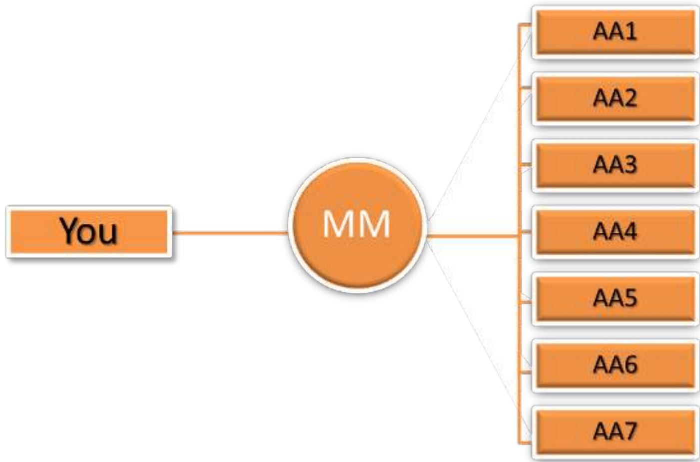
The structure of the market is shown in the following figure. You do not directly trade with other agents in your group, but with MM. That is, if you choose to be a buyer, MM will act as a seller; similarly, if you decide to be a seller, MM will act as a buyer. However, decisions of other members in your group will influence your trading probability, the *Liquidation Value* and your profits.

How to Make Decisions?

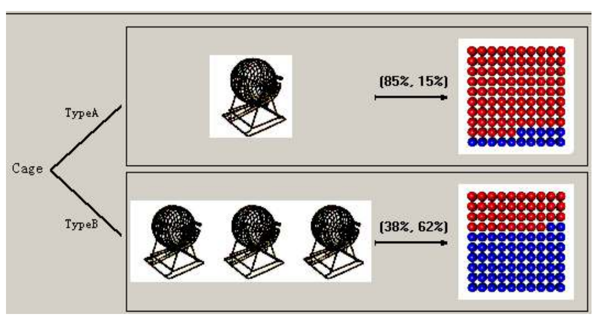
Your decision processes are described in the following steps:

1. **One of four bingo cages has been chosen for your group**

[8] Asset Liquidation (or settlement), is the delivery of an asset by a buyer or a seller.

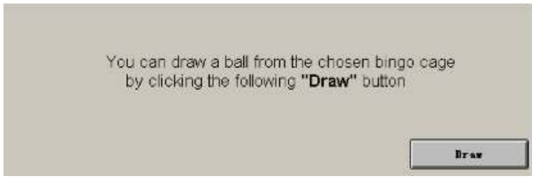


Prior to the start of each round, the computer has already randomly chosen *one* from *four* bingo cages for your group. All of them contain 100 balls of red and blue, but they differ in the proportions of the two colors of balls. For example, in the following figure, a cage of Type A contains 85 red balls and 15 blue balls; three other cages are of Type B and contain 38 red and 62 blue balls. The chosen bingo cage determines the **SAME** probability with which you and your opponents draw a red or a blue ball.



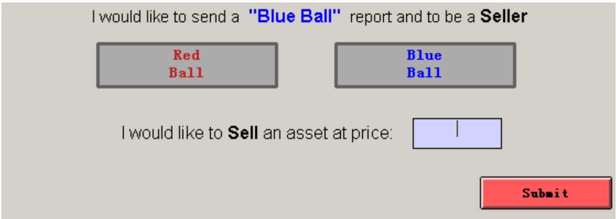
2. **Draw a ball from the chosen bingo cage**

   Without knowing the type of the chosen bingo cage, each agent in your group will draw a ball from it **with replacement**. That means, **every agent in your group has the same probability of drawing a red or a blue ball**. In the following figure, a Blue Ball is drawn from the chosen cage. You will not know any other agent's draw and similarly, your draw will not be revealed to any other agent. However, from your own draw, you can infer which bingo cage could be the chosen one and what are possible draws for your opponents.



3. **Buy or sell an asset on the market**

   After the draw, you need to choose to either buy an asset by sending a "Red Ball" report or sell an asset by sending a "Blue Ball" report. You also need to specify the buying and selling price of the asset, as in the following figure:



What is the Asset on the Market?

The *Market Price* of the asset is randomly drawn from the interval (0, 7) and will be revealed at the end of each round.

The *Liquidation Value* of the asset is determined by your opponents' reports. It is equal to the number of "Red Ball" reports among your 7 opponents.

When does a Trade Occur?

1. **Buyer's Trading Condition**

   If you decide to send a "Red Ball" report, you will automatically become a buyer and you can successfully buy an asset from MM at the *Market Price* if:

   $$\boxed{\textbf{Your Buying Price} \geq \textbf{Market Price} \geq \textbf{MM's Selling Price}},$$

   where **MM's Selling Price** is the average selling price among your opponents.

   The above figure shows a buyer's trading condition for any possible *Market Price* on (0,7). Suppose MS is MM's Selling Price and YB is Your Buying Price, only if the realization of the *Market Price* is between MS and YB, there will be a trade. The associated profit for you is:

   $$\boxed{\textbf{\textit{Buyer's Profit}} = \textbf{\textit{Liquidation Value}} - \textbf{\textit{Market Price}}}$$

2. **Seller's Trading Condition**

   If you decide to send a "Blue Ball" report, you will automatically become a seller and you can successfully sell an asset to MM at the *Market Price* if:

   $$\boxed{\textbf{Your Selling Price} \leq \textbf{Market Price} \leq \textbf{MM's Buying Price}},$$

   where **MM's Buying Price** is the average buying price among your opponents.

   The above figure shows a seller's trading condition. YS is Your Selling Price and MB is MM's Buying Price. There will be a trade for you only if the realization of the *Market Price* falls between YS and MB. The associated profit is:

$$\boxed{\textit{Seller's Profit} = \textit{Market Price} - \textit{Liquidation Value}}$$

3. **Extra Condition**

   When all your opponents choose to buy the asset, No Trade will occur for you. Similarly, No Trade condition applies when all your opponents choose to sell. If there is No Trade, the associated profit is:

$$\boxed{\textit{Profit for No Trade} = \textit{0}}$$

How do Algorithm Agents Make Decisions?

The Algorithm Agents (AAs) are programmed to follow certain decision rules. First, after he draws a ball, he will report the same color of ball as his draw. If he draws a "Red Ball", he will be a buyer by sending a "Red Ball" report, and if he draws a "Blue Ball", he will be a seller by sending a "Red Ball" report. Second, given his draw, he will calculate an expectation of the number of agents who drew a red ball and report it as his buying/selling price.

An Example

Suppose your opponents' decisions are shown in the following table:

Table 8: Opponents' Decisions

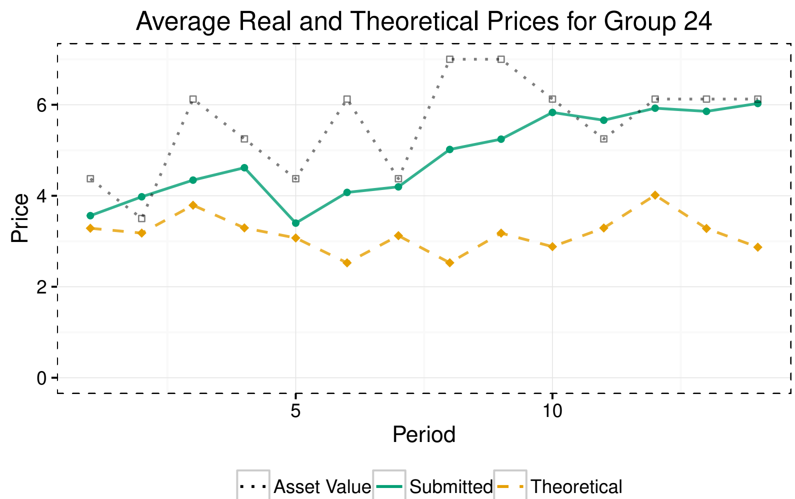
Agent	Draw	Report	Buyer/Seller	Buy/Sell Price
Agent 1	Red Ball	Red Ball	Buyer	4.20
Agent 2	Blue Ball	Blue Ball	Seller	3.50
Agent 3	Blue Ball	Blue Ball	Seller	3.50
Agent 4	Red Ball	Red Ball	Buyer	4.20
Agent 5	Red Ball	Red Ball	Buyer	4.20
Agent 6	Blue Ball	Blue Ball	Seller	3.50
Agent 7	Red Ball	Red Ball	Buyer	4.20

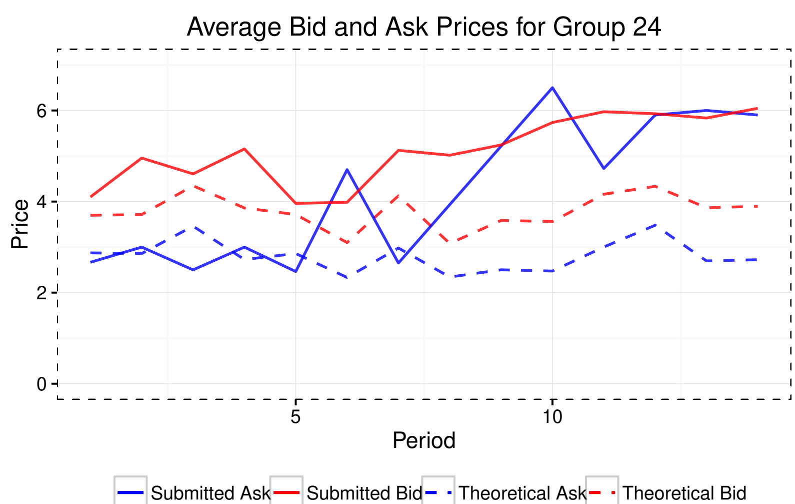
Liquidation Value = number of "Red Ball" reports among your opponents = 4.00.
MM's Buying Price = average buying price among your opponents = 4.20.
MM's Selling Price = average selling price among your opponents = 3.50.

Suppose you want to buy an asset with an offer of 5.00 and the realization of the *Market Price* is 3.80, you can successfully buy an asset because $5.00 \geq 3.80 \geq 3.50$. You need to pay 3.80 and can liquidate this asset at 4.00, thus your profit is $4.00 - 3.80 = 0.20$ and your payoff is 4.20.

Suppose you want to sell an asset at price 3.90 and the realization of the *Market Price* is still 3.80, there will be no trade for you because $3.90 \geq 3.80 \leq 4.20$. If you lower your selling price to 3.30, you will successfully sell an asset to MM. Your profit in this case is $3.80 - 4.00 = -0.20$ and your payoff is 3.80.

Appendix III: Average Price by Groups in 8HA Treatment


Average Real and Theoretical Prices for Group 24
Price
0
2
4
6
5
10
Period
Asset Value
Submitted
Theoretical


Average Bid and Ask Prices for Group 24
Price
0
2
4
6
5
10
Period
Submitted Ask
Submitted Bid
Theoretical Ask
Theoretical Bid

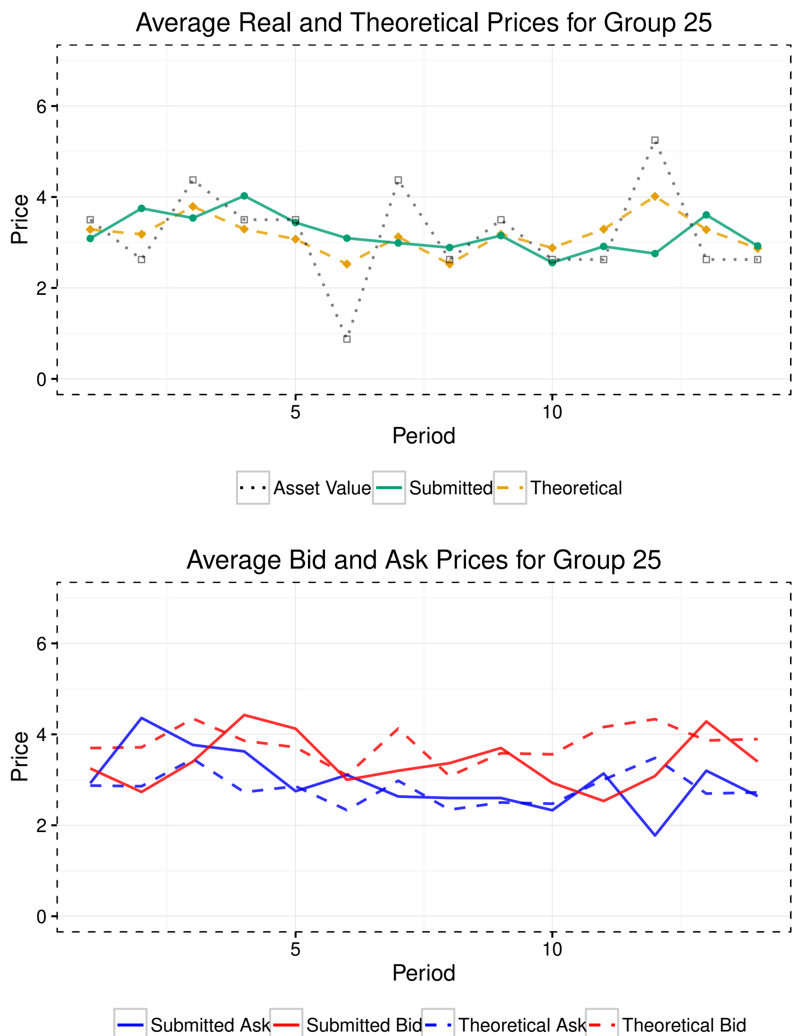


Figure 10: Average Price of Group 25

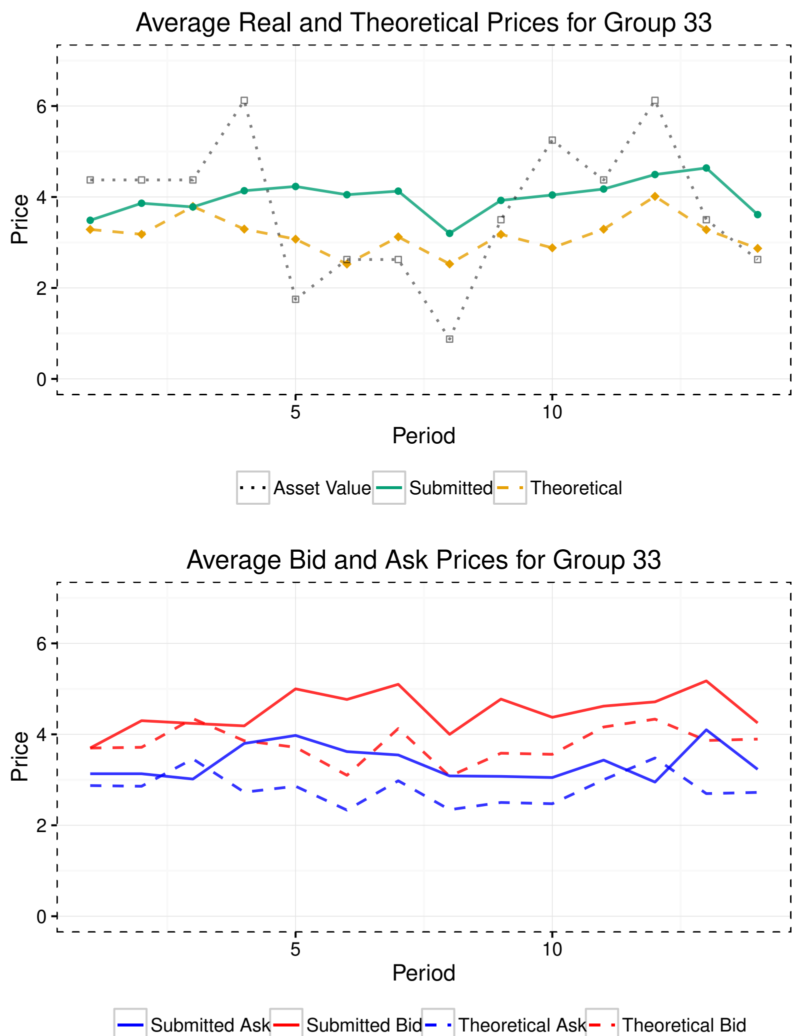
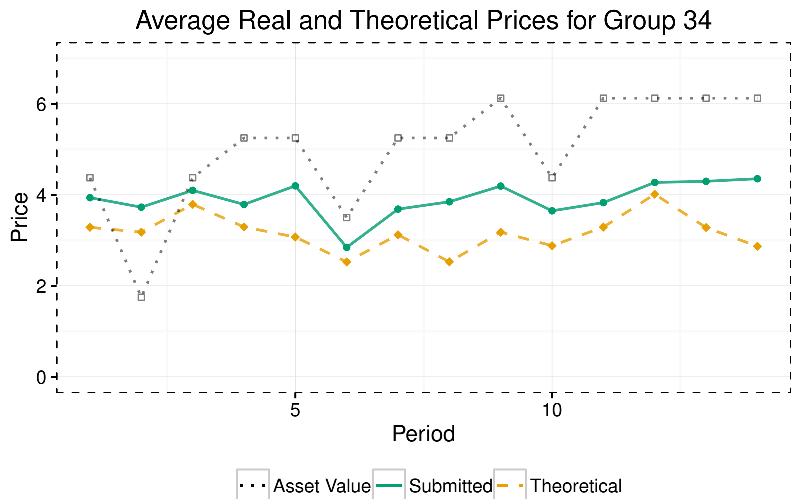


Figure 11: Average Price of Group 33



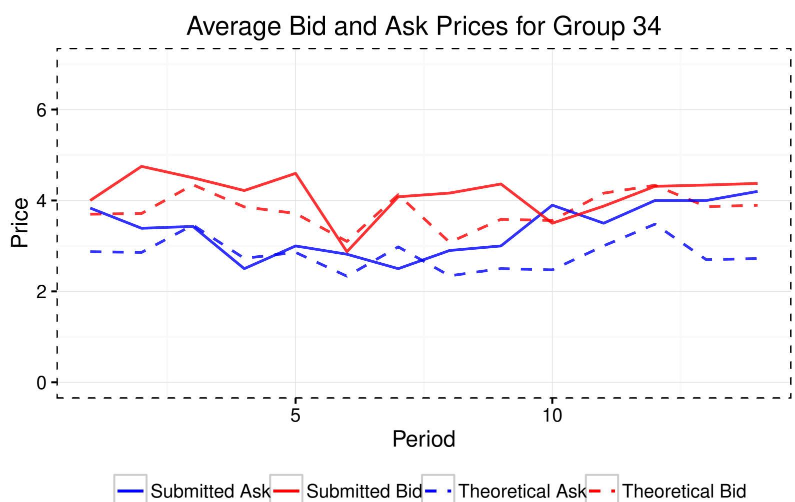


Figure 12: Average Price of Group 34

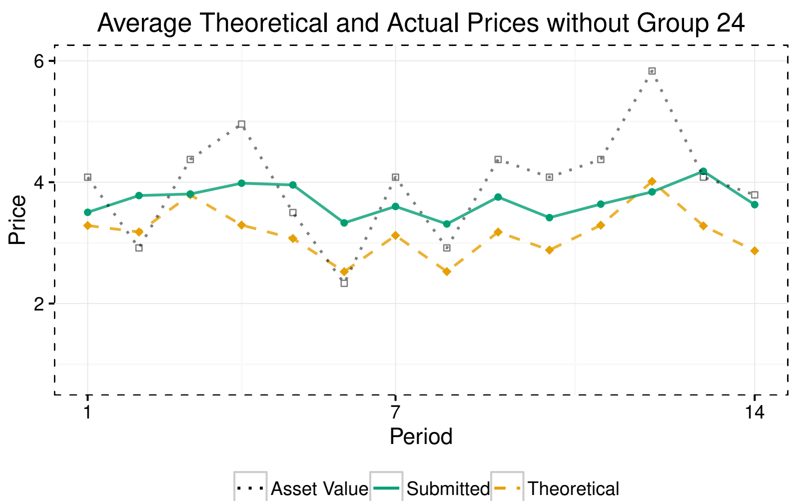


Figure 13: Aggregate Price without Group 24

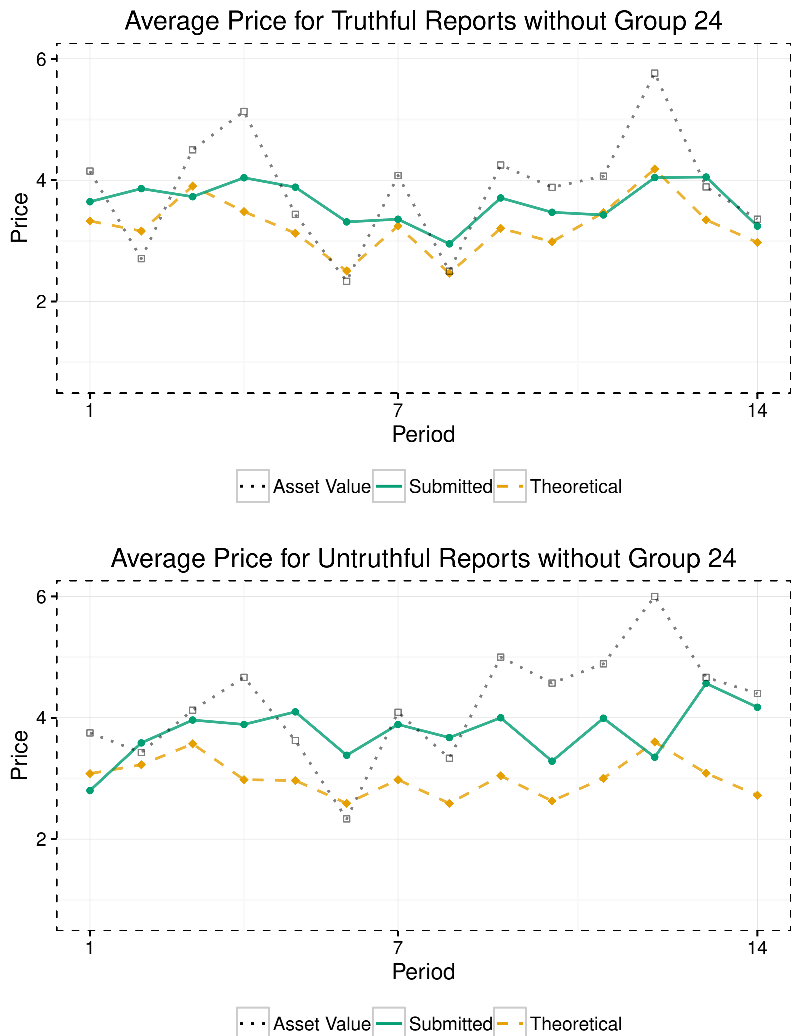


Figure 14: Average Price by Truthfulness without Group 24

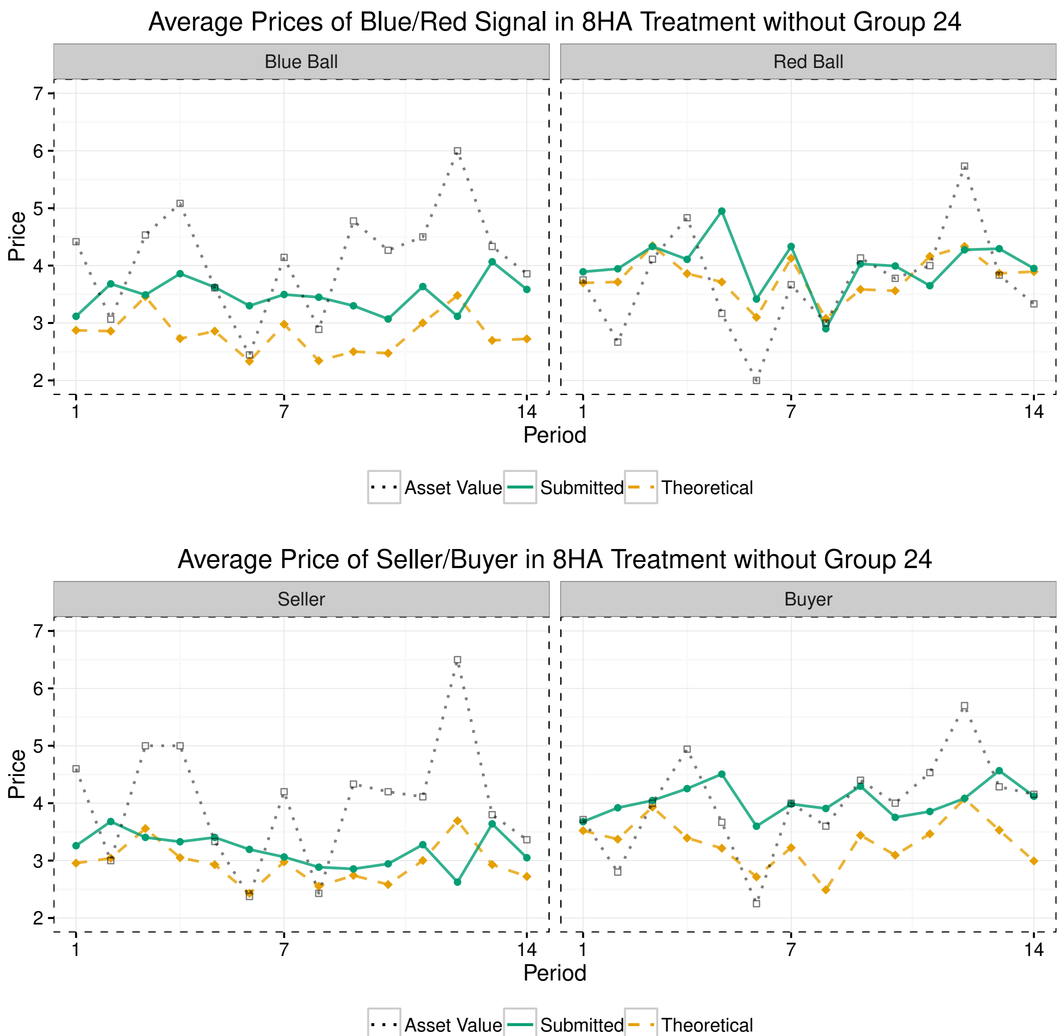


Figure 15: Average Price by Signals and Buyer/Seller without Group 24