Follow the money, not the majority: Incentivizing and aggregating expert opinions with Bayesian markets*

Aurélien Baillon, Benjamin Tereick, Tong V. Wang

Erasmus School of Economics, Erasmus University Rotterdam

August 5, 2020

Abstract

For some questions, such as what the best policy to address a problem is, it is uncertain if the answer will ever be known. Asking experts yields two practical problems: how can their truth-telling be incentivized if the correct answer is unknowable? And if experts disagree, who should be trusted? This paper solves both problems simultaneously. Experts decide whether to endorse a statement and trade an asset whose value depends on the endorsement rate. The respective payoffs of buyers and sellers indicate whom to trust. We demonstrate theoretically and illustrate empirically that "following the money" outperforms selecting the majority opinion.

keywords: wisdom of crowds; expert opinion; truth-telling incentives; prediction markets.

*This research was made possible by European Research Council Starting Grant 638408 BayesianMarkets. We are grateful to Drazen Prelec and Sander Renes for their comments and thank participants of the FinTech and Machine Learning workshop at the National University of Singapore, the 27th SPUDM Conference, the Bayesian Crowd Conference 2019, the Economic Science Association World Meeting 2018, the 18th Conference on the Foundation of Utility and Risk, and seminar audiences at IESEG School of Management, the University of Melbourne, and the University of New South Wales. Email: baillon@ese.eur.nl, tereick@ese.eur.nl, t.wang@ese.eur.nl.

1. Introduction

A centerpiece of economic theory is the idea that markets are efficient aggregators of information (Hayek, 1945; Hurwicz, 1960; Fama, 1970). Historically, this idea has been a descriptive one, explaining the success of market institutions as we encounter them in the real world. In line with an ongoing shift from mere description towards applying economic theory to also create institutions (Roth, 2002, 2018), economists have more recently argued for the use of artificially designed markets ("prediction markets") with information aggregation as a designated goal (Arrow et al., 2008; Hanson, 2003, 2013). Successful applications range from forecasts of political elections (Forsythe et al., 1992; Berg et al., 2008) to business sales (Cowgill and Zitzewitz, 2015; Gillen et al., 2017) and the replicability of experiments in social science (Dreber et al., 2015; Camerer et al., 2016, 2018).

For prediction markets to be successful it is however necessary, both for theoretical and practical reasons, that the true answer to the question they are applied to can be determined within a relatively short time frame. This poses a challenge when we wish to apply them to questions such as what the best policy to address a problem is. For such questions, not only is the answer presently unknown, but it is also uncertain when and how the answer will be known, if at all. If a central bank runs a quantitative-easing policy, we may never be able to assess counterfactual policies (e.g. only using conventional instruments). Furthermore, the best policy may depend on an unobservable state of nature. This creates what we call the "incentive problem": When relying on experts (or on a crowd of laypeople) to provide an answer, how can we incentivize their truth-telling if we do not know whether or when the correct answer will be known?

Besides the incentive problem, we furthermore face an "aggregation problem": which opinion to select if experts disagree? The obvious candidate is the majority opinion, but there is no guarantee that it is the best approach. For instance, imagine each expert can design and run an experiment to test whether a statement is true. Running an experiment can be seen as drawing a binary signal ("support" or "falsify") about the state of nature (whether the statement is true or not). In some extreme cases, a single falsification among many attempts should lead to the rejection of the statement. This would be the case under a strictly Popperian scientific methodology (Popper, 1959) or when validating a mathematical statement, where a single counterexample would be sufficient to establish its falsity. Obviously, in most scientific endeavours, especially in the social sciences, experiments can be noisy and one may expect some experiments falsifying a statement even if it is true. However, the main argument remains: some opinions, based on signals that are more difficult to get, should drive the conclusion even if they are a minority.

In this paper, we study a mechanism which solves both the incentive problem and the aggregation problem simultaneously. We design a market in which experts report their opinions about a statement (endorse it or not) and trade an asset whose value is determined by the total endorsements. Those who endorse the statement are offered to buy the asset from a *center* at price p, where p is randomly drawn. Essentially, buying the asset is betting that more than $p\%$ of others will endorse the statement. Those not endorsing the statement can sell the asset to the center. Baillon (2017) showed that such a "Bayesian market" provides incentives to report opinions truthfully, avoiding the no-trade theorem (Milgrom and Stokey, 1982) through the intermediary role of the center. By making a small adjustment to this mechanism—making the price individualized, independently drawn for each market participant—we show that Bayesian markets have desirable properties with respect to aggregation as well. With sufficiently many participants, experts with the signal that indicates the actual state of nature, and only them, will make a profit. Hence, by "following the money", we can infer the state of nature without relying on what the majority thinks.

The intuition of our result is based on an argument put forward by Prelec et al. (2017). If signals are correlated with the states of nature, there will be more signals supporting a state of nature when this state is the actual one than when it is not, and therefore, than we would have expected ex ante. Prelec et al. (2017) proposed the *surprisingly popular algorithm* (SPA) in which people are asked to endorse a state and predict the rate of endorsement. The algorithm picks the state that is more often endorsed than people predicted. Prelec et al. (2017) demonstrated theoretically and experimentally that this approach improves upon majority and confidence-weighted aggregation.

Bayesian markets allow us to obtain the same improvement but with less information and requiring less cognitive efforts from the participants. We can estimate people's predictions by fitting supply and demand curves for the asset. Furthermore, our method simultaneously provides incentives to truthfully report opinions even if the state of nature is unobservable. Our market approach is not only simpler for participants than the method of Prelec et al. (2017). It also opens up the possibility of continuous markets, extending prediction markets to unverifiable events.

The next section of the paper introduces the theoretical setting and the market. We analyze payoffs at the equilibrium and show how the endorsement of those with positive payoffs indicates the actual state of nature. If the statement is true, those endorsing it can make a profit from betting on others' endorsement rate. If it is not true, those rejecting it can make a profit. The profits realize even in the absence of verification of the actual state of nature, because bets are based on endorsement rates, not on states.

Section 3 describes an experiment we ran on a large sample of US students. We used a task developed by Tereick (2020) that ensures that the informational assumptions of the model are satisfied. Under these assumptions, homo economicus would behave exactly as our model predicts. Our experiment allowed us to test whether our method also worked for homo sapiens, without having to worry whether the informational part of the model perfectly described the reality. We compared our method to the majority opinion and to the SPA. Despite using less information than the SPA, our method had comparable accuracy rates. Both methods substantially improved upon majority when the majority can be wrong.

2. Theory

2.1. Setting

Let $\{Y, N\}$ be the *state space*, with Y and N the two possible *states of nature*. For instance, these two states can represent whether a statement is true or not. Which state S we are in, is assumed to be unobservable.

A group of $n \geq 4$ expert *agents* however has private information about the state.[1] The *common prior* of the agents is that the probability of state Y is r. Each agent gets a *private signal* $s_i \in \{0, 1\}$, with sampling probabilities $P(s_i = 1 \mid Y) = \omega_Y$ and $P(s_i = 1 \mid N) = \omega_N$. Signals are independent conditionally on the state, i.e. $P(s_i = 1 \mid S, s_j) = \omega_S$ for all $S \in \{Y, N\}$ and $j \neq i$.[2] We assume $\omega_Y > \omega_N$. This implies that signals are informative about the state of nature, $s_i = 1$ providing support for Y and $s_i = 0$ for N. We do not require $\omega_Y > 0.5 > \omega_N$, which would be necessary for the majority of signals to be correct (in an infinite group of agents). The assumption $\omega_Y > \omega_N$ is as mild as can be. Equality would mean that s_i is non informative and therefore, all agents would stick to the prior belief r. The opposite inequality would simply change the interpretation of the signal ($s_i = 0$ providing support for Y and $s_i = 1$ for N). Together, we call the triplet $\langle \omega_Y, \omega_N, r \rangle$ a *signal technology*.

Using Bayesian updating, agents form posterior beliefs about the actual state according to

$$r_1 \equiv P(Y \mid s_i = 1) = \frac{r\omega_Y}{r\omega_Y + (1-r)\omega_N}; \quad (1)$$

$$r_0 \equiv P(Y \mid s_i = 0) = \frac{r(1-\omega_Y)}{r(1-\omega_Y) + (1-r)(1-\omega_N)}. \quad (2)$$

For simplicity, we assume that ω_Y, ω_N, and r are such that $r_1 > 0.5$ and $r_0 < 0.5$. It

[1] $n \geq 4$ is required for technical reasons (Baillon, 2017).

[2] In other words, signals are independent and identically distributed given the state, but the latter is uncertain. The absence of correlation between signals implies that agents will not exhibit correlation neglect, unlike studied by Enke and Zimmermann (2019).

allows us to equate an agent's signal with the state the agent believes more likely to be the actual state. If this assumption is not satisfied, signals would be informative but a single signal would not suffice to reverse one's belief. A sufficient condition for this assumption is $r = 0.5$, as used in our experiment.

Apart from the agents' posterior beliefs about states, we can also infer posterior expectations about the proportion of agents who received signal 1 in the population. We denote the actual value of this proportion by ω. Since the expectation of a proportion under random sampling equals the sampling probabilities, agents who received signal 1 expect ω to be

$$\overline{\omega}_1 \equiv E\left[\omega \mid s_i = 1\right] = r_1\, \omega_Y + (1 - r_1)\, \omega_N, \tag{3}$$

whereas agents with signal 0 expect

$$\overline{\omega}_0 \equiv E\left[\omega \mid s_i = 0\right] = r_0\, \omega_Y + (1 - r_0)\, \omega_N. \tag{4}$$

A *center* wants to find out which state we are in (the *actual* state). This center can be a policy maker consulting experts, but could just as well be an employer querying employees or a scientific association surveying its members. We make the usual assumption that the signal technology is common knowledge among the agents. However, as in Prelec (2004), Baillon (2017), Prelec et al. (2017), and Cvitanić et al. (2019), the center does not know the signal technology.

The problem faced by the center is a mechanism design problem, i.e., creating an institution to recover the state of nature given the information structure.[3] Expressed in the terms of our model, the incentive and aggregation problem can be stated as follows. Each agent will report an endorsement e_i, where $e_i = 1$ denotes that agent i endorses state Y and $e_i = 0$ that i endorses state N. The center wants to reward the agents in such a way that it becomes profitable for them to endorse a state if and only if they believe it more likely to be the actual one. Furthermore, upon

[3]By contrast, information design problems keep the payoff structure fix and allow the center to allocate information optimally (Kamenica, 2017, 2019).

learning the endorsements $e_1, \dots, e_n$, the center selects one of the two states, and wishes to maximize the probability that it is the actual one. Since the state S is unobservable and the signal technology is unknown to the center, it is not possible to make the payments or selection of a state dependent on the actual state, nor the selection of the state dependent on the parameters ω_Y and ω_N. Thus, it is impossible to use traditional methods to elicit agents' signals or beliefs because the signals are private (impossible to directly reward truth-telling) and the beliefs are about unverifiable states Y and N (bets and scoring rules cannot be applied). Second, even knowing signals or beliefs would not enable the center to determine the state of nature because the center does not know the values for ω_Y and ω_N. In other words, for anyone unaware of the signal technology, observing 20% of signal 1 does not say which state we are in.

The next subsection introduces the mechanism of our solution concept, called a Bayesian market. Subsection 2.3 presents the underlying idea for an infinitely sized group of expert agents, and addresses the incentive problem and the aggregation problem. It allows us to simplify many practical aspects of the mechanism; for instance, sample proportions and probabilities are equated, with $\omega = \omega_Y$ in state Y and $\omega = \omega_N$ in state N. In Subsection 2.4 we address the case of a finite group of agents. While our results translate to the finite case, our approach still works better the larger the group of experts is.

2.2. Bayesian market

The center and each agent i trade an asset whose *settlement value* v_i is defined as the share of agents other than i endorsing state Y, i.e.,

$$v_i = \frac{\sum_{j \neq i} e_j}{n-1},$$

where excluding agent i's own report prevents agent i from influencing the asset value. The center organizes a *Bayesian market* for these assets:

1. Agents simultaneously report e_i to the center only.

2. For each agent i, the center draws[4] a price p_i from a uniform distribution over $(0, 1)$ and proposes the following trade to the agent, and the agent can decide to take up the offer ($d_i = 1$) or not ($d_i = 0$):

   (a) If $e_i = 1$, agent i can buy the asset at price p_i from the center;

   (b) If $e_i = 0$, agent i can sell the asset at price p_i to the center.

3. All endorsements e_i and buying/selling decisions d_i are revealed.

4. (a) If an agent decides to buy at price p_i, then there is trade under two conditions: (i) There exists another agent j selling at $p_j \leq p_i$ and (ii) there is at least one other agent k who also endorses state Y.

   (b) If an agent decides to sell at price p_i, then there is trade under two conditions: (i) There exists another agent j buying at $p_j \geq p_i$ and (ii) there is at least one other agent k who also endorses state N.

5. Those agents who bought the asset collect v_i and pay p_i; those who sold it collect p_i and pay v_i.

Step 2 differs slightly from the mechanism proposed in Baillon (2017) in which a single price p is drawn for all agents. The motivation for the change is to learn as much as possible from the decisions of different agents. When only a single price is drawn and, e.g., all potential buyers reject the trade, the center only learns that the price was larger than the buyers' reservation price, but not by how much. An alternative would be to directly ask agents for their reservation prices. The center could then draw only one random price p for all agents. This would correspond to the Becker-DeGroot-Marshak mechanism (Becker et al., 1964), but with the trading rule (step 4) in place. The advantage of binary decisions in step 2 is that they require less information from the agents, and therefore less cognitive effort. It is easier to

[4]It is important that the agents are convinced that the prices are independent of their report. To ensure it, the center may draw the prices before step 1, seal them in an envelope, and open it in step 2.

buy/sell at a given price (equivalently, to take/reject a bet on the asset value) than to report a reservation price.[5]

Our mechanism as stated induces a game played among the agents. In this game, a *strategy profile* is a collection $(e, d) = ((e_1, d_1), ..., (e_n, d_n))$, where e_i determines which state individual i is going to endorse depending on the signal s_i, and the trading strategy d_i assigns to each possible signal a range of prices in the $(0, 1)$-interval which i is going to accept when receiving a buy or sell offer from the center. Note that this definition of strategies precludes mixed strategies and the existence of an external coordination device among agents, so that the actual endorsements made by agents are fully determined by their signal and strategy. In Section 5, we discuss this strategy restriction in light of our empirical results.

The mechanism assigns a payoff $U_i(e, d)$ to each agent. Importantly, these payoffs cannot depend on the actual state of nature S or its ω_S. A *Bayesian Nash equilibrium* of the induced game means that, conditioning on their signal, no agent expects a higher payoff by moving to another strategy, i.e.,

$$E\left[U_i\left(e, d\right) \mid s_i\right] \geq E\left[U_i\left(\left(e_1, d_1\right), ..., \left(e_i', d_i'\right), ..., \left(e_n, d_n\right)\right) \mid s_i\right]$$

for any $(e_i', d_i') \neq (e_i, d_i)$ and all signal realizations $s_i \in \{0, 1\}$. We further say that a strategy profile is *truthful* if $e_i(1) = 1$ and $e_i(0) = 0$ for any agent i.

We assume that all agents are risk-neutral[6] and care only about their own monetary payoff, so that $U_i(e, d)$ is just i's monetary payoff. If $e_i = 1$, agent i is potentially a buyer, and we denote by $\pi_1(v_i, p_i)$ agent i's monetary payoff if deciding to buy

[5]Asking for reservation prices, however, has advantages regarding the logistical aspects of practical implementation: In our design, a random price must be drawn for every respondent. When asking for reservation prices, respondents can be contacted by a pen and paper survey in which they submit their reservation prices and a public price is later credibly drawn. Whether these practical considerations outweigh the cognitive simplicity of a binary decision, will depend on the application.

[6]The assumption of risk neutrality is rather common in the literature on expert belief elicitation; see, however, Offerman et al. (2009) and Hossain and Okui (2013) for alternatives.

($d_i = 1$), as a function of the asset value v_i and individualized buying price p_i. Then

$$\pi_1(v_i, p_i) = \begin{cases} v_i - p_i & \text{if trade happens;} \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

Symmetrically, $\pi_0(v_i, p_i)$ denotes agent i's monetary payoff as a potential seller if deciding to sell ($d_i = 0$):

$$\pi_0(v_i, p_i) = \begin{cases} p_i - v_i & \text{if trade happens;} \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

We will first present the intuition of Bayesian markets through the case of an infinite population.

2.3. The case of an infinite population

Three simplifications come with an infinite group of expert agents, which together, imply that the asset value is simply ω_Y or ω_N at the truth-telling equilibrium. First, with n infinite, the proportion of a signal in the population naturally equates the probability to get that signal. Second, excluding the agent's own signal or not from the asset value has no impact, and therefore, the asset individualization becomes equivalent to all agent trading the same asset. The third simplification is related to the trading conditions in step 4 of the Bayesian market definition. That someone else is accepting to buy or to sell at the same price is still important but the information that someone *could* make such a choice becomes trivial in an infinite group. There will always be at least one other experts receiving the same signal and one with the opposite signal. Moreover, for any nondegenerate proportion of agents endorsing each signal, there will also always be someone being offered any possible price. Hence, trade happening does not bring more information about the signal distribution than s_i does, unlike in a finite group of agents. Fortunately, we will see in the next subsection that what there is to learn is negligible when n is finite but sufficiently large.

With the three simplifications in mind, we first address the incentive problem by the following proposition.

Proposition 1. *Let $\langle \omega_Y, \omega_N, r \rangle$ be a signal technology and n infinite. In the game induced by the Bayesian market, truth-telling is a Bayesian Nash equilibrium in which agents' betting strategies are such that:*
(i) agents whose signal is 1 buy the asset if and only if $p_i \leq \bar{\omega}_1$;
(ii) agents whose signal is 0 sell the asset if and only if $p_i \geq \bar{\omega}_0$.

Proof. The main result in Baillon (2017) is essentially unaffected by the introduction of individualized prices. To get an intuition for the result, we can inspect Equations (3) and (4). It is immediate that $\overline{\omega}_0 < \overline{\omega}_1$ since $r_0 < r_1$ and $\omega_N < \omega_Y$. Thus, signal-1 agents expect more signal-1 agents than signal-0 agents do. Consider then agent i with $s_i = 1$ and assume all other agents are telling the truth, such that the asset value v_i equals the true share of signal-1 agents in the population. Agent i expects v_i to be $\overline{\omega}_1$. For p_i less than $\overline{\omega}_1$, agent i will be willing to buy the asset. Agent i also knows that no one would buy it at a higher price (so i has no reason to pretend to be a seller) but that some agents will be willing to sell at prices between $\overline{\omega}_0$ and $\overline{\omega}_1$. For this price range, agent i foresees a profit and has the incentives to endorse $e_i = 1$ to become a buyer. Outside this range, no trade will go through. The case $s_i = 0$ is symmetric. □

The fact that agents trade an asset whose value they disagree on may raise the question why the no-trade theorem (Milgrom and Stokey, 1982) is not applicable here. The reason lies in the role of the center: For a trade to go through, it is sufficient that there exists a single agent who was willing to take the opposite bet. The center will verify this condition for each individual bettor, without providing further information about who the agent with the opposite bet is. Since $0 < \omega_N < \omega_Y < 1$, agents already know that there must be at least one disagreeing agent and thus the occurence of trade does not provide further information about the actual

ω. Since trades are facilitated by the center,[7] the agents remain uncertain about the share of other agents disagreeing with them, which makes our setting different to the settings in Aumann (1976) or Milgrom and Stokey (1982) in which disagreement is impossible.

In the following proposition, we consider the aggregation problem and derive what conclusions the center can draw in the truth-telling equilibrium.

Proposition 2. *If n is infinite and the Bayesian market is at the truth-telling equilibrium, at least one agent has a positive payoff, all those with positive payoffs have endorsed the actual state, and all those with negative payoffs have endorsed the opposite state.*

Proof. At the truth-telling equilibrium, the settlement value v_i is ω_N in state N and ω_Y in state Y. And according to Proposition 1, trades only occur for prices in the range $[\overline{\omega}_0, \overline{\omega}_1]$. Hence agents' payoffs, defined in Equations (5)-(6), can be simplified as

$$\pi_1(v_i, p_i) = -\pi_0(v_i, p_i) = \begin{cases} \omega - p_i & \text{if } p_i \in [\overline{\omega}_0, \overline{\omega}_1]; \\ 0 & \text{otherwise.} \end{cases}$$

Notice that Equations (3)-(4) imply

$$0 < \omega_N < \overline{\omega}_0 < \overline{\omega}_1 < \omega_Y < 1. \tag{7}$$

In state Y, when trade occurs, signal-1 agents pay less than $\overline{\omega}_1$ and therefore less than the settlement value ω_Y. They make a profit while sellers (signal-0 agents) sell the asset at a price too low. The opposite applies in state N. Hence, the center, seeing that people endorsing Y make a profit, can conclude that we are indeed in state Y, even though the state itself is not directly observable. Sellers making a profit indicates state N. □

[7]The center will typically incur a loss from this role. The mechanism is thus not budget-balanced.

In this infinite case, at the truth-telling equilibrium and under the actual state of nature S, the average payoff for agents with the same signal s is equal to the expected payoff for agents with that signal:

$$\begin{aligned}
\pi_1^Y &\equiv E_p[\pi_1(v_i, p) \mid Y] \\
&= E_p[\pi_1(\omega_Y, p)] = \int_{\overline{\omega}_0}^{\overline{\omega}_1} (\omega_Y - p) dp = \frac{1}{2}\left[(\omega_Y - \overline{\omega}_0)^2 - (\omega_Y - \overline{\omega}_1)^2\right] \qquad (8) \\
&= -\pi_0^Y \equiv -E_p[\pi_0(v_i, p) \mid Y];
\end{aligned}$$

$$\begin{aligned}
\pi_0^N &\equiv E_p[\pi_0(v_i, p) \mid N] \\
&= E_p[\pi_0(\omega_N, p)] = \int_{\overline{\omega}_0}^{\overline{\omega}_1} (p - \omega_N) dp = \frac{1}{2}\left[(\overline{\omega}_1 - \omega_N)^2 - (\overline{\omega}_0 - \omega_N)^2\right] \qquad (9) \\
&= -\pi_1^N \equiv -E_p[\pi_1(v_i, p) \mid N].
\end{aligned}$$

Under state Y, $\pi_1^Y > 0$ and $\pi_0^Y < 0$; and under state N, $\pi_1^N < 0$ and $\pi_0^N > 0$.

The value $\overline{\omega}_s$ is the prediction of the proportion of signal 1 in the population by agents with signal s. Hence, $\omega_S - \overline{\omega}_s$ is the prediction error of signal-s agents when S is the actual state of nature (note that this error can be positive or negative). The average payoff of signal-s agents are therefore half the difference between the squared prediction error of agents with signal $1-s$ and their own squared prediction error. Agents endorsing the actual state of nature are better able to guess the signal distribution in the population, and therefore, the opinions of others. Bayesian markets favor them and allow them to make a profit.

2.4. Adjustments to a finite population

We presented the results with n infinite. However, with a small group of agents, the individualized prices may affect the equilibrium strategies. For a buyer, the existence of an agent being offered to sell at a specific price is informative about the number of sellers, and therefore informative about the asset value. Agents may then accept to buy or sell at prices that do not give rise to the aggregation property pointed out by Proposition 2.

Intuitively, the information contained in the existence of an agent on the opposite side of the market should decrease with the number of agents. In Appendix A, we make this intuition rigorous: a "sufficiently large" group of agents restores properties used by Propositions 1 and 2. Specifically, we show that for any signal technology there is a truth-telling equilibrium in which agents endorsing the actual state make a positive payoff with arbitrarily large probability, given that n is larger than some finite threshold n^*. Thus, as the number of market participants increases, the center can be almost certain that the agents making money are the ones who correctly identify the actual state.

2.5. Algorithms for empirical data

Propositions 2 concerns limit behavior of perfectly rational agents. In perfect conditions, all agents endorsing the actual (opposite) state have a nonnegative (nonpositive) payoff, and at least one agent will have a positive payoff. In practical implementation, a small group may lead to no trade. Furthermore, agents may make mistakes when endorsing a state or when deciding to trade. In the presence of noise, agents not endorsing the actual state may still make a profit. We propose two *follow-the-money (FTM)* algorithms which can be used empirically by the center to find the actual state in those non-ideal situations.

The simpler algorithm computes the payoff of each agent and compares the average payoff of the sellers to that of the buyers. The algorithm picks the side with the higher average payoff and tosses a coin if no trade occurred. We call this algorithm *FTM-A* (for average). FTM-A is able to accommodate some moderate noise in agents' behavior but does not solve the no-trade issue.

To account for noise and for no-trade situations, we propose a more elaborate algorithm, fitting logistic curves, called *FTM-L*. With F the logistic function,[8] FTM-L first estimates $\hat{\omega}_1$ and $\hat{\omega}_0$ (which can be interpreted as the reservation prices for

[8] If F is the probit function, we can define another algorithm called *FTM-P*. All the reported results are robust to this specification. We will focus on FTM-L from now on.

an infinite group at the truth-telling equilibrium) from

$$Prob(d_i = 1 \,|\, p, e_i) = \begin{cases} F(\beta(p - \hat{\omega}_1)) & \text{if } e_i = 1 \\ F(\beta(\hat{\omega}_0 - p)) & \text{if } e_i = 0 \end{cases} \tag{10}$$

imposing $\hat{\omega}_0 \leq \hat{\omega}_1$. Parameter β captures the level of noise/imprecision and is assumed to be the same for sellers and buyers (for parsimony). FTM-L then computes the expected payoffs for buyers and sellers for an infinite group using Equations (8) or (9), substituting $\bar{\omega}_1$ and $\bar{\omega}_0$ with estimated reservation prices $\hat{\omega}_1$ and $\hat{\omega}_0$, and ω with the proportion of endorsements 1, and picks the side with a positive expected payoff.

3. Experimental design

3.1. Stimuli

We conducted an experiment with abstract tasks (urns and balls) ensuring that the theoretical assumptions were satisfied. We considered groups of $n = 100$ agents. In each task, the participants of the experiment were presented with two urns, as depicted in Figure 1. Urns Left and Right represent the two states of nature, N and Y respectively. Participants were told that one of the two urns was selected randomly ($r = 0.5$) and that each of the 100 participants of a group would get one ball from that urn. Denoting a yellow ball $s_i = 1$ and a blue ball $s_i = 0$, Urn Left would give $\omega_N = 0.10$ and Urn Right $\omega_Y = 0.40$ in this particular example. Urn Right always contains more yellow balls than Urn Left. Thus Urn Right is state of nature Y and Urn Left is state of nature N.

There were 30 tasks with ω_N ranging from 0.05 to 0.75 and ω_Y from 0.25 to 0.95, spanning the unit interval in a systematic way. In twelve tasks, both urns had a minority of yellow balls, i.e., $\omega_N < \omega_Y < 0.5$. Another set of twelve tasks mirrored them such that $\omega_Y > \omega_N > 0.5$, and in six tasks the majority would always guess the correct state of nature ($\omega_Y > 0.5 > \omega_N$). Table 3 in Online Appendix III.1 lists

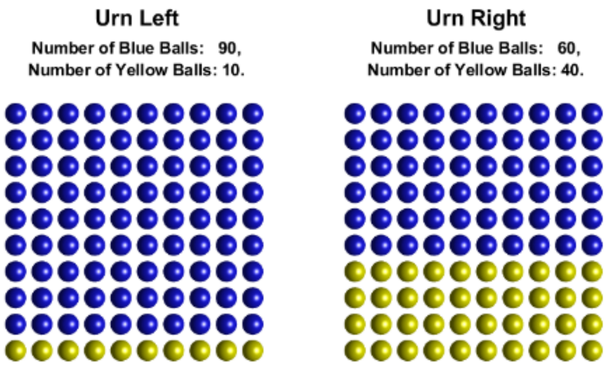


Figure 1: Experimental task setting (an example of Task 6)

all the task parameters. The number of yellow balls differs across states of nature by a minimum of 20 and maximum of 30. Larger differences would mean that the signal technology discriminates very well between state of nature and the majority (as well as FTM) would be right most of the time. By contrast, smaller differences would imply very narrow trading intervals $[\overline{\omega}_0, \overline{\omega}_1]$ and it could be that none of the 100 participants of a group gets a price in that range.

In each task, the participants were presented with the urns (as in Figure 1) and asked to press a button to draw their ball. Once the color of their ball was revealed, they were asked to guess which urn the ball comes from (i.e. to endorse a state). The next question differed between two experimental treatments, FTM and SPA.

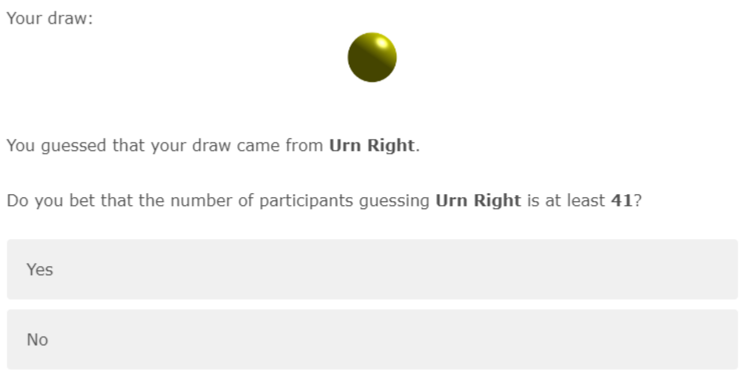


Figure 2: Screenshot of the FTM treatment

In the FTM treatment, we implemented the betting mechanism of the Bayesian markets. In Figure 2 for instance, participants were asked whether they were willing to bet that the number of participants guessing Urn Right (i.e. endorsing Y) was at least 41, i.e. whether they were willing to pay $p = 0.41$ for the sample proportion ω. For the sake of symmetry, participants guessing Urn Left were asked whether they would bet that the number of participants guessing Urn Left would be at least 59, i.e. whether they were willing to accept $p = 0.41$ for v. Payment was explained in a training preceding the experiment. The participants were told that the number (e.g. 41) was random and that their payment would be the actual number of Urn Right guesses minus that number if someone took the opposite bet (betting that at least 59 participants would guess Urn Left). It would be 0 otherwise.

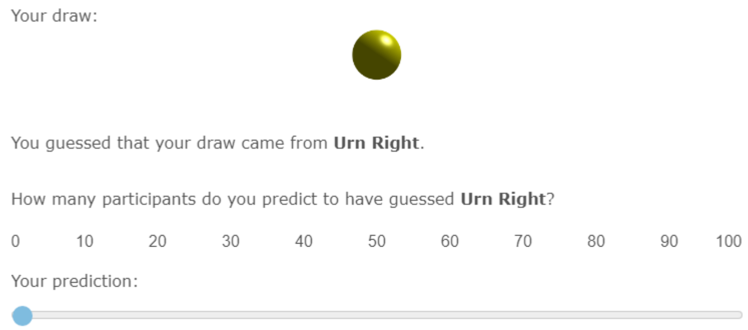


Figure 3: Screenshot of SPA.

In the SPA treatment, we followed the approach of Prelec et al. (2017) and asked participants to predict the number of people who guessed the same urn as they did (Figure 3). Prelec et al.'s (2017) algorithm first computes the average prediction across all participants and then select the state of nature that was endorsed more often than predicted. Predictions were incentivized using the quadratic scoring rule. Participants received $200 - \frac{x^2}{50}$, with x the difference between their prediction and the actual number of guesses.

In none of the treatments was payment directly based on the task parameters. Even though we, the experimenters, knew them, we aimed to mimic situations in which no one knows the actual state of nature and in which the center (paying the agents) does not even know the signal technology.

3.2. Deviations from Section 2.2

The implementation of the Bayesian market in our experiment differs from the Bayesian market mechanism proposed in subsection 2.2 in two ways. First, the settlement value v did not exclude the agent's own endorsement. It allowed us to present all relevant values as shares of 100 (and not 99 or 101). Furthermore, the draws from the urn (i.e., the signals) were made without replacement. As a consequence, the settlement value could only be ω_Y or ω_N, as in the infinite case. In Online Appendix III.1, we show that these design choices do not affect the equilibrium properties of the market for the parameters of the experiment. They only simplify calculations for respondents. The online appendix also shows that 100 market participants is sufficiently large for conditioning on the occurrence of trade to be negligible. Thus, the incentives and aggregation properties of the Bayesian market are essentially as explained in Section 2.3 where n is infinite.

3.3. Implementation

The experiment was conducted on Prolific in August 2019, with 473 participants in the FTM treatment and 462 in the SPA treatment. They were all US students. We restricted participation to students for their probable familiarity with abstract tasks as those used in our experiment. Participants watched a short video explaining the experimental tasks and then went through five training rounds where they received feedback about their payments and how these payments were calculated (see Online Appendix III for details). We split the 30 tasks into two sets of 15. After the training, each participant completed one of the two sets, with the task order being randomized within that set at the participant level. There was no feedback after the tasks. Payment, described in the next paragraph, occurred once all participants had completed the experiment.

Participants received a fixed reward of £1.5 and a bonus of up to £3.[9] All

[9]Prolific required payments in pounds.

amounts (prices, bets, scores) were presented in tokens. The bonus in pounds was the number of tokens divided by 1,000. In the FTM treatment, participants could (in theory) win or lose up to 100 tokens in each task. Hence, they were endowed with 100 tokens for each task to avoid net losses at the end of the experiment. In the SPA treatment, the quadratic score was also expressed in tokens. It was equivalent to endowing them with 200 tokens and imposing a quadratic loss ranging from 0 to 200. In both treatments, the final number of tokens was naturally bounded by 0 and 3000. This allowed us to recruit participants with the same information about bonus ranges. However, the average bonus was likely to be lower for the FTM treatment than for the SPA treatment ex ante and, in fact, it was ex post (SPA £2.85, FTM £1.60).

To compute the bonus of a participant in a given task after the end of the experiment, we randomly selected a state of nature[10] and 100 participants such that the group (including this particular participant) had the exact combination of signals shown in the task. In other words, participants were not assigned to a given group ex ante. Instead, we constructed (random) groups matching the information provided to the participants.

4. Results

To be consistent, we report data and results in terms of our theoretical setting. In particular, a yellow ball is signal 1 ($s_i = 1$) and a blue ball is signal 0 ($s_i = 0$). A participant guessing Urn Right is endorsing state of nature Y ($e_i = 1$) and guessing Urn Left is endorsing N ($e_i = 0$).[11] We also define truth-telling as reporting $e_i = s_i$.

[10]This random selection of a state of nature resulted in 50.2% of state Y selected for bonus calculations of participants in the SPA treatment, and 48.9% for the FTM treatment. Both proportions are not significantly different from 0.5 (proportion tests: for SPA, Z-statistic$= 0.333$, $p = 0.739$; and for FTM, Z-statistic$= 1.904$, $p = 0.057$).

[11]Predictions elicited in the SPA treatment were about the number of participants guessing the same urn, but we deduct the predictions of participants endorsing N from 100 to be the predictions

4.1. Raw data - Endorsements

According to the model, truth-telling is a Bayesian Nash equilibrium in the FTM treatment. The empirical truth-telling rate was 87.6%.[12] About 53% of the participants told the truth in all 15 tasks they faced. About 25% guessed the opposite urn (or lied about their guess) 1 to 3 times out of 15. Less than 4% had a majority of lies / wrong guesses (Table 4 in Online Appendix III.2). The incentives provided in the SPA treatment did not make truth-telling a Bayesian Nash equilibrium, but we observed a very similar truth-telling rate (SPA: 87.8% of the cases, not significantly different from FTM, with proportion test Z-statistic$= 0.376$ and $p = 0.704$).

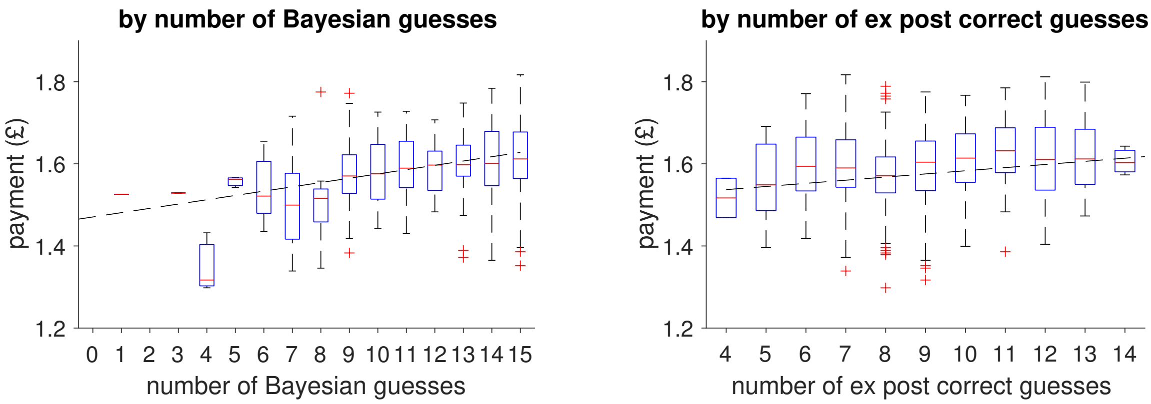


Figure 4: Payments in FTM.

The truth-telling rate of the FTM treatment was sufficiently high to reward those who correctly reported their signal and to penalize those who lied or misreported. The left panel of Figure 4 displays earnings as a function of the number of times people told the truth. It shows a positive correlation, with a fitted line slope of

of number of participants endorsing Y. Bets in the FTM treatment were also expressed in terms of the number of participants guessing the same urn, but we deduct the prices in the bets for participants endorsing N from 100 to be the the prices to sell the asset whose settlement value is the number of participant endorsing Y.

[12]The empirical truth-telling rates were not significantly different for easier questions with $\omega_Y > 0.5 > \omega_N$ and for other questions with $\omega_Y > \omega_N > 0.5$ or $0.5 > \omega_Y > \omega_N$ (87.0% and 87.7% respectively; proportion test Z-statistic$= 0.755$ and $p = 0.450$).

0.010 ($p < 0.001$). People did not get feedback during the experiment (only in the five training rounds). The figure illustrates that feedback about payment could have improved truth-telling rate by allowing participants to learn that correctly reporting their signal is rewarded. It further shows that in future experiments one can announce in the instructions that a previous study showed that participants who tell the truth more often can earn more in such a setting.

So far, we studied what the raw data told us about participants' strategic behavior, illustrating the incentive properties of Bayesian markets (Proposition 1). To illustrate the aggregation properties (Proposition 2), we can check whether correctly guessing the selected urn led to higher earnings in our experiment. The prediction is supported by the right panel of Figure 4, which is a box plot of earnings as a function of the number of times participants guessed the actual state. The fitted line slope is 0.008 ($p < 0.001$). Thus, Bayesian markets reward expertise. While in our experiment, this expertise is artificially created,[13] in many applications one may expect that the number of times someone guesses the actual state of the world to be influenced by a more natural notion of expertise, i.e. domain knowledge.

4.2. Raw data - Predictions and trades

If participants are Bayesian, they should report the posteriors $\overline{\omega}_0$ and $\overline{\omega}_1$ in the SPA treatment, at least if they expect everyone else to tell the truth. Figure 5 displays the average predictions as a function of theoretical posteriors for both type of guess. Predictions are very close to Bayesianism for $\overline{\omega}_0 < 0.5$ and $\overline{\omega}_1 > 0.5$. Interestingly, participants seemed to have much more difficulty to predict that a majority of people would guess Y when they themselves guess N or that only a minority would guess Y when they themselves guess Y. The SPA uses the average prediction across both guesses, which mitigates this issue.

We do not have people's predictions in the FTM treatment but we can compare the participants' decisions d_i to the theoretical predictions. Table 1 compares the

[13]It consists of receiving informative signals, in combination with a truth-telling strategy.

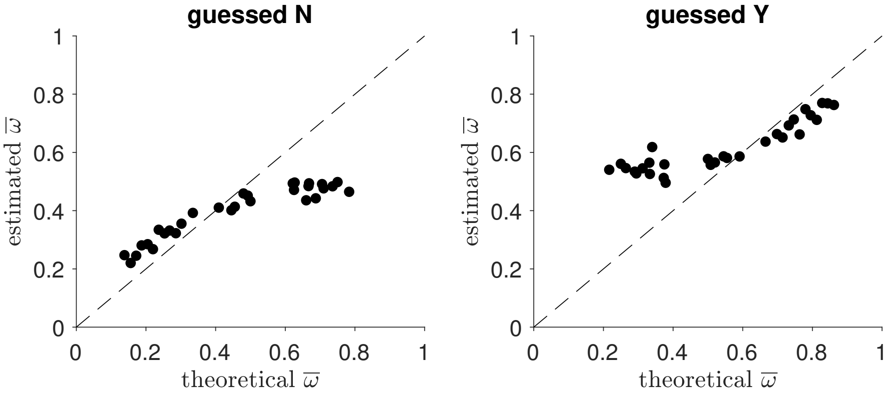


Figure 5: Theoretical $\overline{\omega}_0$ and $\overline{\omega}_1$ vs. average predictions in SPA.

theoretical and empirical proportions of $d_i = 1$ (the willingness to buy / to sell) for five price intervals, defined with ω_N, $\overline{\omega}_0$, $\overline{\omega}_1$, and ω_Y. Buyers should be willing to pay at most $\overline{\omega}_1$ and sellers willing to accept not less than $\overline{\omega}_0$. If participants do not compute the Bayesian posterior but use ω_Y and ω_N instead, i.e. the distribution of balls of the urn they guessed, buyers would be willing to pay at most ω_Y and sellers willing to accept not less than ω_N. If they were extremely risk averse, buyers would be willing to pay at most ω_N and sellers willing to accept not less than ω_Y.

Table 1: Theoretical and empirical bet acceptance (in %) and average payoffs (in tokens) in the FTM treatment by price interval

	$p \in$	$[0, \omega_N)$	$[\omega_N, \overline{\omega}_0)$	$[\overline{\omega}_0, \overline{\omega}_1]$	$(\overline{\omega}_1, \omega_Y]$	$(\omega_Y, 1]$
guessed N (seller)	theo. acceptance	0%	0%	100%	100%	100%
	emp. acceptance	30.4%	55.5%	67.6%	80.8%	90.6%
	average payoff	-27.2	-6.1	2.2	10.8	19.3
guessed Y (buyer)	theo. acceptance	100%	100%	100%	0%	0%
	emp. acceptance	89.9%	80.3%	67.7%	48.4%	29.1%
	average payoff	21.0	9.9	1.3	-6.9	-25.0

The empirical willingness to sell was increasing with price and the empirical willingness to buy was decreasing, as predicted in the truth-telling equilibrium.

However, for several participants the acceptance and rejection ranges of bets were not consistent with the equilibrium prediction. About 30% of bets that are losing for sure under truth-telling were accepted and about 10% of bets that are winning for sure under truth-telling were rejected (see leftmost column of the seller row and rightmost column of the buyer row). In total, there was a clear tendency to bet much more than predicted by equilibrium play.

Table 1 also reports the average payoffs of the participants for each price interval. The results confirm that participants who accepted bets that would have been losing for sure if everyone else had told the truth, still bore a loss on average in our experiment. Overall, trading decisions were noisy and substantially deviated from the theoretical predictions. This underlines that the performance of the FTM algorithms will depend on their ability to recover aggregate reservation prices from the noisy trades.

4.3. Accuracy comparison

The final part of the analysis aims to compare accuracy of the various methods. We want to assess the ability of the majority rule, SPA, and FTM algorithms to identify the actual state of nature using the participants' answers.

To make full use of the answers of all respondents who provided answers to a task, we ran 1,000 simulations for each task, state of nature, and treatment, randomly making groups of 100 participants. For instance, consider one of the simulations for the task described in Figure 1 with $\omega_Y = 0.40$ and $\omega_N = 0.10$ (Task 6), state of nature Y, and the FTM treatment. We randomly composed a group of 100 FTM participants, such that exactly 40 of them had gotten $s_i = 1$. We then use the answers from the 100 participants to determine the state using majority rule, FTM-A, and FTM-L. Similarly, we randomly composed 1,000 groups of 100 SPA participants in the same way to determine the state using majority rule and SPA. Repeating the same procedures for each of the 30 tasks and two possible states of nature, we obtained 60 accuracy rates for each method. Table 2 summarizes the average accuracy rates for each algorithm and for the majority rule. We conducted Wilcoxon tests to test for differences.

Table 2: Average accuracy rates from simulations

cluster of questions	majority rule		SPA	FTM	
	data SPA	data FTM		*FTM-A*	*FTM-L*
$\omega_Y > 0.5 > \omega_N$	95.7%	99.7%	93.2%	87.3%	91.5%
$\omega_Y > \omega_N > 0.5$ or $0.5 > \omega_Y > \omega_N$	51.0%	53.8%	73.9%	62.8%	75.0%

Table 2 distinguishes two cases. If $\omega_N < 0.5 < \omega_Y$ (top row), then the majority rule should determine the actual state all the time. In the other cases (bottom row), the majority rule finds the actual state 50% of the time, by pure chance. Our results are consistent with these predictions (see columns 'majority rule'), both for the data from the SPA treatment and for the data of the FTM treatment. The SPA, our benchmark, should always identify the actual state if participants were Bayesian and reporting truthfully all the time. Non-Bayesian answers and noise make the SPA perform worse (Wilcoxon signed rank test $p = 0.031$) than majority when $\omega_N < 0.5 < \omega_Y$ but substantially improved upon majority when following the majority is equivalent to tossing a coin (Wilcoxon signed rank test $p < 0.001$). In that case, the average accuracy increased by 22.9 percentage points (pp).

Computing average payoffs on Bayesian markets, as our FTM-A algorithm does, led to worse results than SPA, whether $\omega_N < 0.5 < \omega_Y$ (Wilcoxon signed rank test $p = 0.054$) or not (Wilcoxon signed rank test $p = 0.001$). FTM-A is highly sensitive to noise and we noticed earlier that our data were clearly noisy. To account for noise, FTM-L fits logistic supply and demand curves on the buy and sell decisions and only then computes expected payoffs. FTM-L substantially improved upon FTM-A (Wilcoxon signed rank test $p < 0.001$), especially when $\omega_Y > \omega_N > 0.5$ or $0.5 > \omega_Y > \omega_N$, with an increase of 12.2pp. It yielded results that were not significantly different from SPA (Wilcoxon signed rank tests; top row $p = 0.322$ and bottom row $p = 0.845$). Interestingly, it gave results comparable to SPA with less information. SPA uses, as input, an endorsement and a prediction (number between 0 and 1), directly asking participants for $\overline{\omega}_0$ and $\overline{\omega}_1$. FTM-L uses an endorsement and a trade decision, which is binary. FTM-L compensates the information loss by using (simple) econometric techniques to recover reservation prices, which should be $\overline{\omega}_0$ and $\overline{\omega}_1$ at the truth-telling equilibrium.

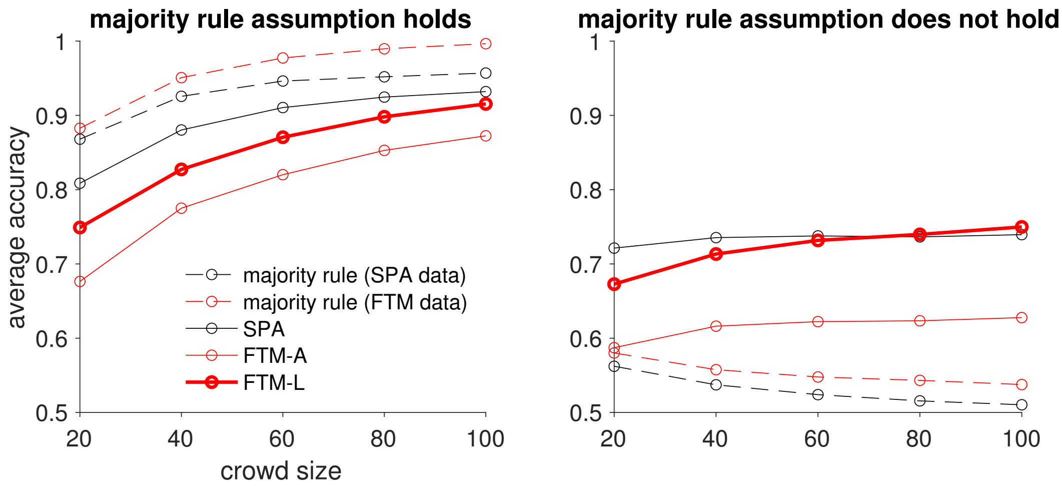


Figure 6: Accuracy comparisons for different group sizes.

The accuracy analysis so far was based on groups of 100 participants. We can also study how sensitive results are to group sizes. We replicated the analysis (with 1,000 simulations for each combination of method, task, and state of nature) for various group sizes ranging from 20 to 100. Figure 6 depicts the accuracy rates as a function of group size. FTM-L is more sensitive to group size than SPA. In the left panel, when $\omega_N < 0.5 < \omega_Y$, accuracy of FTM-L increases from around 75% for groups of 20 to more than 90% for groups of 100. SPA performs better for small groups but accuracy increases less with group size. In the right panel ($\omega_Y > \omega_N > 0.5$ or $0.5 > \omega_Y > \omega_N$), SPA is very stable, with accuracy rates between 72% and 74%. The accuracy of FTM-L is lower than that of SPA for groups of 20 but slightly higher for groups of 100. These results are not surprising, knowing that, for groups of 20, FTM-L has to determine reservations prices from very few binary decisions (buying or selling).[14]

5. Discussion

In our experiment, each agent received an endowment to avoid losses. Even without providing an endowment, agents can expect a strictly positive payoff (Baillon, 2017), which can motivate them to participate. The center, who plays the role of the market maker, subsidizes the market and acts as an intermediary between the agents, who do not trade with each other. Absent this intermediary role of the center, agents would infer others' signals from their willingness to buy or sell. Similar to the classical reasoning in Aumann (1976) and Milgrom and Stokey (1982), they would then agree on the state, leaving no room for trade based on disagreement.

To communicate the same information to all potential participants, we fixed the bonus range from £0 to £3. The SPA was more expensive (SPA £2.85, FTM £1.60). If anything, the SPA participants, with an endowment of 200 tokens and a quadratic loss should have been more motivated than the FTM participants. The SPA treatment only incentivized predictions, not truthful endorsement. The latter could have been done using the Bayesian truth serum of Prelec (2004) but the payoff rule is difficult to explain to participants. Experiments that have been using this truth serum did not explain the payoff function in detail, but rather used an "intimidation method", telling participants it is in their interest to tell the truth. We refrained from such an approach, and instead included instructions and training to explain our payoff rules. An alternative for future research is to incentivize the SPA using choice-matching (Cvitanić et al., 2019), which elicits predictions and endorsements with a simpler payment formula.

Regarding our theoretical model, four restrictive assumptions warrant some further discussion. First, recall that that we treated strategies as maps from signals

[14]There may be as little as one buyer or one seller in some groups even if everybody reports truthfully.

to endorsements, such that agents could not make their endorsements depend on any other event or randomization devise, and did not allow asymmetric strategies. Second, we allowed no communication between agents. Third, we only considered a binary underlying state space and fourth, our market setting is a one-period, static setting. We discuss each of those in turn.

It is important that agents cannot coordinate on events other than type realizations. Among the remaining symmetric equilibria, the truth-telling equilibrium is (ex-ante) Pareto optimal.[15] It is behaviorally plausible, as conjectured by Baillon (2017), that truth-telling is focal and, in our experiment, there was indeed little evidence of agents trying to find a reverse strategy. More than half of the participants consistently told the truth and a negligible share of participants chose to systematically misstate their type (see Table 4 in Online Appendix III.2). Without the aforementioned restriction, agents could try to coordinate on some other signal, in which the probability of receiving a 0-signal in state Y and a 1-signal in state N is very low. Then, a small number of agents will make a loss of (almost) 1, and a large share of agents will make a profit of (almost) 1. In expectation, all agents thus have a high expected payoff. Note that this coordination does not only require mere communication among respondents but also some credible randomization device. To avoid such coordinated attacks, the center should make it an active feature of design that market participants are (at least partially) anonymous, as is the case in our experiment.

As suggested by the previous paragraph, our approach cannot be used if there is public discussion of private signals or if agents can form coalitions. If it is possible to bring all experts together, other approaches to the aggregation problem have been proposed in the literature, such as the Delphi method developed in the 1950s at the Rand Corporation (Okoli and Pawlowski, 2004). These approaches do not solve the

[15] To see this, note first that it is pay-off equivalent to the "reverse" equilibrium in which everyone endorses the state that they believe to be less likely to be the actual one. The two other equilibria have universal endorsement of either state Y or of state N and thus obviously lead to a universal payoff of zero since there is never any trade.

incentive problem though. In our setting, experts do not have other incentives than those we provide to hide or manipulate their private information. The literature on committee decisions studies how agents may agree to share their private signals with each other in order to look united if their reputation is at stake (Visser and Swank, 2007; Swank et al., 2008).

We considered a binary state space. If the state space is non-binary, one may organize several Bayesian markets, with different agents. Consider three states A, B, and C, and assume agents can choose which state they would like a signal about (e.g., an agent can design an experiment testing whether we are in state A or not-A). The center can assign agents to markets, inform them about which state their market will be about, let them run their experiment for that state, and then organize the Bayesian markets.

Bayesian markets and their aggregation properties can further be translated to a setting in which a market is run continuously. Suppose that there are T periods and that for each $t = 1, ..., T$, a Bayesian market is set up to trade on an asset v_t that represents the share of buyers in the Bayesian market at time t. All of these markets are only settled at the final period T, so that in particular agents do not learn the value of the assets. At each t, the incentive and aggregation properties of Bayesian markets are not affected by the markets in other periods. A continuous market can sometimes be advantageous for the center: Suppose for instance that the signal technology is constant across all periods, but that the actual state S (and therefore ω_S) may vary with t. Once the center has found a market-clearing price p^* (i.e. a price at which each agent is willing to either buy or sell the asset), this price can be chosen for any subsequent period. Since the signal technology is the same, this price will now lead to trade in each period, thereby reducing the payoff-uncertainty faced by the agents. Then, the center can make inferences about the change of the state over time by computing which side would make a profit if the market was settled. Furthermore, if the signal technology is not fixed, this will be reflected in the buying and selling decisions of the agents, and henceforth the center can detect such changes.

The literature on the wisdom of crowds started with the intuition that asking many people may be better than relying on a few experts. Some have raised doubts on the mere possibility to “chase the experts” within a group (Larrick and Soll, 2006). However, there is still value to ask large groups of experts. DellaVigna and Pope (2017) found that the aggregated opinion of academic experts is closer to experimental results than estimates based on a meta-analysis of previous empirical findings. In a follow-up study, DellaVigna and Pope (2018) also showed that academic experts better predict than non-experts, even though degrees of expertise (among experts) such as academic rank or citations do not correlate with performance. Aggregating the opinions of very large group of experts becomes more and more common, for instance the International Panel on Climate Change or surveys of economists and financial specialists about future economic indicators.

6. Conclusion

Prediction markets are increasingly used to incentivize and aggregate expert opinions. They are not applicable though if the state of the world is not objectively observable. In such a case, payoffs cannot be state-contingent, creating an incentive problem. Furthermore, in many plausible situations, one may prefer not to rely on the majority opinion, at least if experts themselves, aware of the signal structure, would not. We demonstrated theoretically and empirically how to solve both the incentive and the aggregation problem at once. Agents bet on others' endorsement and their payoffs reveal the state of nature. When implemented in a large online experiment, our follow-the-money approach performed as well as a recent alternative, the surprisingly popular algorithm, with less information from participants.

Increasingly in companies, prediction markets are used internally among employees to forecast short-term company performance and external events for decision making. Examples include Siemens (Ortner, 1998), Nokia (Hankins and Lee, 2011), Hewlett-Packard (Plott and Chen, 2002), Intel (Gillen et al., 2017), Google (Cowgill et al., 2009), and Ford Motor Company (Cowgill and Zitzewitz, 2015). Results are promising, showing the potential of markets as an effective information aggregation tool in practice. For instance, in the case of Ford where weekly auto sales forecasts are taken extremely seriously for planning procurement and production, forecasts from the internal prediction markets still outperformed other forecasts available to management (Cowgill and Zitzewitz, 2015). Since Bayesian markets do not require the predicted events to be verifiable in the short-term or at all, they expand the horizon of prediction markets to long-term events or even to counterfactual and unverifiable events.

A. Appendix - Results for the finite case

Proposition 3. *Let $\langle \omega_Y, \omega_N, r \rangle$ be a signal technology. Then: (i) truth-telling is an equilibrium in the game induced by the Bayesian Market, and (ii) in this equilibrium, for any $\varepsilon > 0$, there exists n^* such that for all $n \geq n^*$, the probability that at least one agent has a positive payoff, and all those with positive payoffs have endorsed the actual state, and all those with negative payoffs have endorsed the opposite state, is at least $1 - \varepsilon$.*

To simplify the notations in the following proofs, we define $\omega_{-i} = \frac{\sum_{j \neq i} s_j}{n-1}$, $\tilde{P}_s(.) \equiv P\left(. \,|\, s_i = s, \omega_{-i} \in (0,1)\right)$ and $\tilde{E}_s[.] \equiv E\left[. \,|\, s_i = s, \omega_{-i} \in (0,1)\right]$, suppressing the conditioning on i's signal being s and the fact that there is at least one signal-1 agent and one signal-0 agent other than i. Furthermore, $\tilde{r}_s \equiv \tilde{P}_s(Y)$ and $\tilde{\omega}_s \equiv \tilde{E}_s\left[\omega_{-i}\right]$. For any agent i with price p_i, let $\mathcal{E}_0^{-i}(p_i)$ be the event that there exists another agent j such that $s_j = 0$ and $p_j \leq p_i$, and $\mathcal{E}_1^{-i}(p_i)$ the event that there exists another agent j such that $s_j = 1$ and $p_j \geq p_i$.

All the proofs in the following subsections and in the Online Appendices establish the results for $s_i = 1$. Equivalent proofs for $s_i = 0$ can be immediately obtained by replacing s_i, e_i, ω_{-i}, p_i by $1 - s_i$, $1 - e_i$, $1 - \omega_{-i}$, $1 - p_i$. Recall that $n \geq 4$ for all the proofs.

A.1. Preparatory lemmas

The following lemmas are used to establish Proposition 3 (Lemma 1 being used to establish Lemma 5). The lemmas are not especially surprising, and the proofs are more cumbersome than truly informative. We therefore relegated them to Online Appendix I.

Lemma 1. $\lim_{n\to\infty} \tilde{\omega}_1 = \bar{\omega}_1$ *and* $\lim_{n\to\infty} \tilde{\omega}_0 = \bar{\omega}_0$.

Lemma 2. *For any $\varepsilon > 0$, there exists n^* such that for all $n \geq n^*$, $P(\omega_{-i} > \bar{\omega}_1 \,|\, Y) > 1-\varepsilon$, $P(\omega_{-i} < \bar{\omega}_0 \,|\, N) > 1-\varepsilon$, $P(\omega_{-i} > \tilde{\omega}_1 \,|\, Y) > 1-\varepsilon$, and $P(\omega_{-i} < \tilde{\omega}_0 \,|\, N) > 1-\varepsilon$.*

Lemma 3. *$\tilde{E}_1\left[\omega_{-i}\,|\,\mathcal{E}_0^{-i}(p_i)\right] < \tilde{\omega}_1 < \tilde{E}_1\left[\omega_{-i}\,|\,\mathcal{E}_1^{-i}(p_i)\right]$ and $\tilde{E}_0\left[\omega_{-i}\,|\,\mathcal{E}_0^{-i}(p_i)\right] < \tilde{\omega}_0 < \tilde{E}_0\left[\omega_{-i}\,|\,\mathcal{E}_1^{-i}(p_i)\right]$.*

Lemma 4. *$\tilde{E}_1\left[\omega_{-i}\,|\,\mathcal{E}_0^{-i}(p_i)\right]$ and $\tilde{E}_0\left[\omega_{-i}\,|\,\mathcal{E}_1^{-i}(p_i)\right]$ are both continuous in p_i.*

Lemma 5. *$\lim_{n\to\infty}\tilde{E}_1\left[\omega_{-i}\,|\,\mathcal{E}_0^{-i}(p_i)\right] = \bar{\omega}_1$ and $\lim_{n\to\infty}\tilde{E}_0\left[\omega_{-i}\,|\,\mathcal{E}_1^{-i}(p_i)\right] = \bar{\omega}_0$.*

A.2. Proof of Proposition 3 - Part (i)

Suppose that all agents are truth-telling and further that signal-1 agents will buy at any price p for which $p \leq \tilde{E}_1\left[\omega_{-i}\,|\,\mathcal{E}_0^{-i}(p)\right]$ and signal-0 agents will sell at any price p for which $p \geq \tilde{E}_0\left[\omega_{-i}\,|\,\mathcal{E}_1^{-i}(p)\right]$. To see that these strategies constitute an equilibrium, we take the perspective of a buyer, and it is completely analogous to show that there is no profitable deviation for a seller. Consider an agent i who observes signal $s_i = 1$ and price p_i. Under the strategy described above, i never accepts to trade at a price for which i expects a negative payoff and thus i's expected payoff is strictly positive. Any deviation in which i reports $e_i = 1$ but buys at a price $p_i > \tilde{E}_1\left[\omega_{-i}\,|\,\mathcal{E}_0^{-i}(p_i)\right]$ is at best non-harmful, if no seller would sell at this price in equilibrium, or leads to a lower expected payoff. Suppose then next that i deviates by reporting $e_i = 0$. Afterwards, i cannot do better than selling at any price p_i for which $p_i > \tilde{E}_1\left[\omega_{-i}\,|\,\mathcal{E}_1^{-i}(p_i)\right]$. From Lemma 3, we know that $\tilde{E}_1\left[\omega_{-i}\,|\,\mathcal{E}_1^{-i}(p_i)\right] > \tilde{E}_1\left[\omega_{-i}\,|\,\mathcal{E}_0^{-i}(p_i)\right]$. Hence, all other signal 1-agents reject such trades, and therefore, i's payoff will be zero. Thus, there is no profitable deviation for i and the strategies described above indeed constitute an equilibrium.

A.3. Proof of Proposition 3 - Part (ii)

To prove the second part of Proposition 3, we need to find a group size such that we are almost certain that at least one agent has a positive payoff, and all those with positive payoffs have endorsed the actual state, and all those with negative payoffs have endorsed the opposite state. Alternatively, we can show that, for a given group size and above, the probability of the following three events is negligible:

- Event 1: there exists an agent endorsing the actual state making a loss;
- Event 2: there exists an agent endorsing the other state making a profit;
- Event 3: no agent trades.

Let us first show that Event 2 implies Event 1. In state Y, an agent endorsing $e_i = 0$ can only make a profit if there is a buyer j with $p_j \geq p_i > v_i$. Since v_i excludes $e_i = 0$ and v_j excludes $e_j = 1$, we have $v_j = v_i - \frac{1}{n-1} < v_i$. Then $p_j > v_j$, implying that agent j must be making a loss. The same argument can be made for state N. Consequently, Event 2 is a subevent of Event 1 and can be ignored.

We can establish $P($"agents endorsing the actual state, and only them, will make a profit"$) > 1 - \varepsilon$ for all n above a threshold by showing that $P(\text{Event 1}) < \varepsilon/2$ and $P(\text{Event 3}) < \varepsilon/2$. Recall that statement (ii) posits that we are in the truth-telling equilibrium. We can therefore equate v_i with ω_{-i}.

Event 1: there exists an agent endorsing the actual state making a loss. From Lemma 2, it follows that we can find n_1 such that in the truth-telling equilibrium with probability at most $\frac{\varepsilon}{2}$, we have $v_i \leq \tilde{\omega}_1$ in state Y or $v_i \geq \tilde{\omega}_0$ in state N. Furthermore, from Lemma 3, we have that for any price p and any n: $\tilde{E}_1 \left[v_i \,|\, \mathcal{E}_0^{-i}(p) \right] < \tilde{\omega}_1$, and $\tilde{E}_0 \left[v_i \,|\, \mathcal{E}_1^{-i}(p) \right] > \tilde{\omega}_0$. Thus, in the equilibrium described in part (i) of Proposition 3, there will never be trade for prices exceeding $\tilde{\omega}_1$ (because buyers will not accept such prices) and for prices lower than $\tilde{\omega}_0$ (because sellers will not accept such prices). Combining the two results, we know that there exists an n_1 such that for all $n \geq n_1$, the probability that any trade occurs at a price larger than the asset value in state Y (implying that some buyers – who endorse the correct state – make a loss) is at most $\frac{\varepsilon}{2}$, and analogously for state N.

Event 3: no agent trades. To obtain the probability of no trade, we will first derive the probability of trade. Trades occur if there exists an interval of prices $\mathcal{I}$ for which people are willing to trade and prices are drawn in $\mathcal{I}$ for sellers and buyers.

Willingness to trade: Let $p_m = \frac{\bar{\omega}_0+\bar{\omega}_1}{2}$ (the "midpoint" between $\bar{\omega}_0$ and $\bar{\omega}_1$). Choosing $\delta = \frac{\bar{\omega}_1-\bar{\omega}_0}{3}$, Lemma 5 implies that there exists n_δ such that $\bar{\omega}_1 - \tilde{E}_1\left[v_i \,|\, \mathcal{E}_0^{-i}(p_m)\right] < \delta$ and $\tilde{E}_0\left[v_i \,|\, \mathcal{E}_1^{-i}(p_m)\right] - \bar{\omega}_0 < \delta$ for all $n \geq n_\delta$. Hence, $\tilde{E}_1\left[v_i \,|\, \mathcal{E}_0^{-i}(p_m)\right] > \bar{\omega}_1 - \delta > p_m$ and $\tilde{E}_0\left[v_i \,|\, \mathcal{E}_1^{-i}(p_m)\right] < \bar{\omega}_0 + \delta < p_m$, which show that both buyers and sellers are willing to trade at price p_m, according to the proof of (i). Furthermore, by continuity of $\tilde{E}_1\left[v_i \,|\, \mathcal{E}_0^{-i}(p)\right]$ in p (Lemma 4), there must be an interval $\mathcal{I} = [p_m - a,\, p_m + a]$ such that $\tilde{E}_1\left[v_i \,|\, \mathcal{E}_0^{-i}(p)\right] > p$ for all $p \in \mathcal{I}$ (and analogously for $\tilde{E}_0$). We have therefore shown that for all $n \geq n_\delta$, there is an interval of prices $\mathcal{I}$ for which people are willing to trade.

Prices being drawn in $\mathcal{I}$: In what follows, we keep assuming $n \geq n_\delta$. Let $\mathcal{E}^+$ be the event that there is at least one buyer with $p_i \in [p_m, p_m + a]$ and $\mathcal{E}^-$ be the event that there is at least one seller with $p_i \in [p_m - a, p_m]$. Then $P(\mathcal{E}^+ \,|\, k \text{ buyers}) = 1 - (1-a)^k$ and $P(\mathcal{E}^- \,|\, n-k \text{ sellers}) = 1 - (1-a)^{n-k}$. Hence, under any state of the world $S \in \{Y, N\}$,

$$
\begin{aligned}
& P(\text{trade in } \mathcal{I} \,|\, S) \\
\geq & P\left(\mathcal{E}^+ \text{ and } \mathcal{E}^- \,|\, S\right) \\
= & \sum_{k=1}^{n-1} P(k \text{ buyers and } n-k \text{ sellers} \,|\, S)\, P\left(\mathcal{E}^+ \text{ and } \mathcal{E}^- \,|\, k \text{ buyers and } n-k \text{ sellers}\right) \\
= & \sum_{k=1}^{n-1} P(k \text{ buyers and } n-k \text{ sellers} \,|\, S)\, P\left(\mathcal{E}^+ \,|\, k \text{ buyers}\right) P\left(\mathcal{E}^- \,|\, k \text{ sellers}\right) \\
= & \sum_{k=1}^{n-1} \binom{n}{k} \omega_S^k (1-\omega_S)^{n-k} \left[1 - (1-a)^k\right]\left[1 - (1-a)^{n-k}\right] \\
= & 1 + (1-a)^n - \left[(1-a)\omega_S + (1-\omega_S)\right]^n - \left[\omega_S + (1-a)(1-\omega_S)\right]^n .
\end{aligned}
$$

Since the limit of the last expression is 1,

$$
\lim_{n\to\infty} P(\text{no trade in } \mathcal{I}) = 1 - \lim_{n\to\infty} P(Y)P(\text{trade in } \mathcal{I} \,|\, Y) + P(N)P(\text{trade in } \mathcal{I} \,|\, N) = 0.
$$

This implies that there exists an $n_2 \geq n_\delta$ such that for all $n \geq n_2$, there is at most $\frac{\varepsilon}{2}$ probability that no trade occurs in the interval $\mathcal{I}$. Note that n_2 must be at least n_δ to ensure the existence of $\mathcal{I}$ in which agents are willing to trade.

Final step. The probability that at least one agent has a positive payoff, and all those with positive payoffs have endorsed the actual state, and all those with negative payoffs have endorsed the opposite state, is at least $1 - P(\text{Event } 1) - P(\text{Event } 3)$. For all $n \geq n_1$, the probability of Event 1 is at most $\frac{\varepsilon}{2}$ and for all $n \geq n_2$, that of Event 3 is also at most $\frac{\varepsilon}{2}$. Let $n^* = \max\{n_1, n_2\}$. Then for any $n \geq n^*$, the probability that some agents endorsing the actual state, and only them, will make a profit and no agent endorsing the actual state makes a loss is more than $1 - \varepsilon$, as desired.

A.4. Simulations

Proposition 3 is a limit result and we conducted simulations to estimate what a sufficient group size should be. For $r = 0.5$, all pairs $(\omega_N, \omega_Y) \in \{0.05, 0.10, \ldots, 0.95\}$ with $\omega_N < \omega_Y$, and various group sizes, we checked how often agents endorsing the actual state and only them make a profit. The results of our simulations assume that all pairs (ω_N, ω_Y) described above are equally likely, even though situations in which ω_N and ω_Y are very close corresponds to signal technologies that are only very weakly informative.[16] Online Appendix II describes the simulation procedure and the left panel of Figure 7 displays the results.

It is worth noting that situations in which agents endorsing the wrong state are making a profit are extremely unlikely. Moreover cases in which some agents of both sides make a loss are basically non-existing. Thus, whenever there is trade, which side has a positive payoff reveals the actual state with very high accuracy, even for small group sizes. However, for small numbers of agents ($n < 100$), the most likely outcome is that no agent trades, and therefore, no one makes a profit. This still happens in about 25% of the cases for large group sizes (e.g. 300).[17] There are two

[16]For example, if $\omega_Y - \omega_N = 0.05$ and $r = 0.5$, the average accuracy of the signal technology is 52.5%, to be compared with 50% of random guessing.

[17]The reason is that in more than one third of our simulations, the difference between $\bar{\omega}_0$ and $\bar{\omega}_1$ is less than 0.05. If there are only few agents receiving a given signal, it becomes unlikely that one of them will also be offered a price in the interval in which trades could occur.

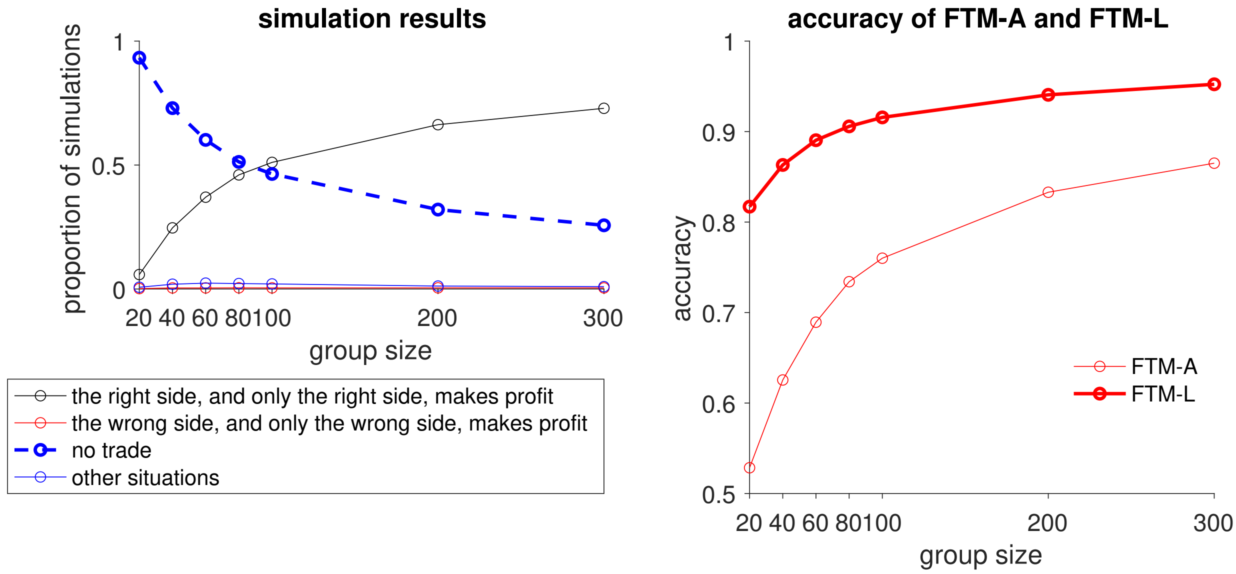


Figure 7: Simulations for different group sizes n.

main ways to address this issue. First, any information about the signal technology (for instance a lower bound on ω_N or an upper bound on ω_Y) would allow the center to draw prices in a more specific region instead of uniformly between 0 an 1, thereby increasing the probability of trade. Second, the absence of trades does not mean that no information was collected. From the agents' endorsements and trade decisions, the center can recover information using our two algorithms to infer the actual state.

The right panel of Figure 7 displays the accuracy, defined as the number of times the actual state is selected, of both FTM-A and FTM-L algorithms in the same simulations as described in the preceding paragraph. The simulations do not include noisy agents and therefore, FTM-A cannot improve much with respect to requiring that all agents make a profit to trust their endorsement. By contrast, FTM-L reaches an 80% accuracy with groups as small as 20 agents. It reaches a 95% accuracy with 300 participants, which is not an unusual sample size in social sciences. The cases in which FTM-L is inaccurate correspond to signal technologies that are only very weakly informative. In all other cases, FTM-L recovers all necessary information even in the absence of trades.

References

Arrow, K. J., Forsythe, R., Gorham, M., Hahn, R., Hanson, R., Ledyard, J. O., Levmore, S., Litan, R., Milgrom, P., Nelson, F. D., Neumann, G. R., Ottaviani, M., Schelling, T. C., Shiller, R. J., Smith, V. L., Snowberg, E., Sunstein, C. R., Tetlock, P. C., Tetlock, P. E., Varian, H. R., Wolfers, J., and Zitzewitz, E. (2008). The promise of prediction markets. *Science*, 320(5878):877–878.

Aumann, R. J. (1976). Agreeing to disagree. *The Annals of Statistics*, pages 1236–1239.

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.

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

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.

Cowgill, B., Wolfers, J., and Zitzewitz, E. (2009). Using prediction markets to track information flows: Evidence from google. In *AMMA*, page 3.

Cowgill, B. and Zitzewitz, E. (2015). Corporate prediction markets: Evidence from google, ford, and firm x. *The Review of Economic Studies*, 82(4):1309–1341.

Cvitanić, J., Prelec, D., Riley, B., and Tereick, B. (2019). Honesty via choice-matching. *American Economic Review: Insights*, 1(2):179–92.

DellaVigna, S. and Pope, D. (2017). What motivates effort? evidence and expert forecasts. *The Review of Economic Studies*, 85(2):1029–1069.

DellaVigna, S. and Pope, D. (2018). Predicting experimental results: Who knows what? *Journal of Political Economy*, 126(6):2410–2456.

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.

Enke, B. and Zimmermann, F. (2019). Correlation neglect in belief formation. *The Review of Economic Studies*, 86(1):313–332.

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

Forsythe, R., Nelson, F., Neumann, G. R., and Wright, J. (1992). Anatomy of an experimental political stock market. *The American Economic Review*, 82(5):1142–1161.

Gillen, B. J., Plott, C. R., and Shum, M. (2017). A pari-mutuel-like mechanism for information aggregation: A field test inside intel. *Journal of Political Economy*, 125(4):1075–1099.

Hankins, R. and Lee, A. (2011). Crowd sourcing and prediction markets. Working paper.

Hanson, R. (2003). Combinatorial information market design. *Information Systems Frontiers*, 5(1):107–119.

Hanson, R. (2013). Shall we vote on values, but bet on beliefs? *Journal of Political Philosophy*, 21(2):151–178.

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

Hossain, T. and Okui, R. (2013). The binarized scoring rule. *Review of Economic Studies*, 80(3):984–1001.

Hurwicz, L. (1960). Decentralization and computation in resource allocation. In Arrow, K. J., editor, *Mathematical methods in the social sciences*. Stanford University Press.

Kamenica, E. (2017). Information economics. *Journal of Political Economy*, 125(6):1885–1890.

Kamenica, E. (2019). Bayesian persuasion and information design. *Annual Review of Economics*, 11:249–272.

Larrick, R. P. and Soll, J. B. (2006). Intuitions about combining opinions: Misappreciation of the averaging principle. *Management science*, 52(1):111–127.

Milgrom, P. and Stokey, N. (1982). Information, trade and common knowledge. *Journal of Economic Theory*, 26(1):17–27.

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.

Okoli, C. and Pawlowski, S. D. (2004). The delphi method as a research tool: an example, design considerations and applications. *Information & Management*, 42(1):15–29.

Ortner, G. (1998). Forecasting markets: An industrial application. Technical report, mimeo.

Plott, C. R. and Chen, K.-Y. (2002). Information aggregation mechanisms: Concept, design and implementation for a sales forecasting problem.

Popper, K. (1959). *The Logic of Scientific Discovery*. Routledge.

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

Prelec, D., Seung, H. S., and McCoy, J. (2017). A solution to the single-question crowd wisdom problem. *Nature*, 541(7638):532–535.

Roth, A. E. (2002). The economist as engineer: Game theory, experimentation, and computation as tools for design economics. *Econometrica*, 70(4):1341–1378.

Roth, A. E. (2018). Marketplaces, markets, and market design. *American Economic Review*, 108(7):1609–58.

Swank, J., Swank, O. H., and Visser, B. (2008). How committees of experts interact with the outside world: some theory, and evidence from the fomc. *Journal of the European Economic Association*, 6(2-3):478–486.

Tereick, B. (2020). Improving information aggregation through meta-cognition. Working paper.

Visser, B. and Swank, O. H. (2007). On committees of experts. *The Quarterly Journal of Economics*, 122(1):337–372.

For Online Publication

Online Appendix I: Proof of the lemmas

I.1. Proof of Lemma 1

Proof. Using that $\tilde{P}_1(\omega_{-i}=0)=\tilde{P}_1(\omega_{-i}=1)=0$, since $\tilde{P}$ conditions on the fact that $\omega_{-i}\in(0,1)$, and further using that

$$\tilde{P}_1\left(\omega_{-i}=\frac{k}{n-1}\right)=\frac{P\left(\omega_{-i}=\frac{k}{n-1}\,|\,s_i=1\right)}{1-P\left(\omega_{-i}=0 \text{ or } \omega_{-i}=1\,|\,s_i=1\right)}$$

for $0<k<n-1$, we get

$$\begin{aligned}
\tilde{\omega}_1=&\sum_{k=1}^{n-2}\frac{k}{n-1}\tilde{P}_1\left(\omega_{-i}=\frac{k}{n-1}\right)\\
=&\frac{1}{1-P\left(\omega_{-i}=0 \text{ or } \omega_{-i}=1\,|\,s_i=1\right)}\sum_{k=1}^{n-2}\frac{k}{n-1}P\left(\omega_{-i}=\frac{k}{n-1}\,|\,s_i=1\right)\\
=&\frac{1}{1-P\left(\omega_{-i}=0 \text{ or } \omega_{-i}=1\,|\,s_i=1\right)}\left[E(\omega_{-i}\,|\,s_i=1)-P\left(\omega_{-i}=1\,|\,s_i=1\right)\right]\\
=&\frac{1}{1-P\left(\omega_{-i}=0 \text{ or } \omega_{-i}=1\,|\,s_i=1\right)}\left[\frac{n}{n-1}\bar{\omega}_1-\frac{1}{n-1}-P\left(\omega_{-i}=1\,|\,s_i=1\right)\right].
\end{aligned}$$

Note that in the last step, we use $\omega_{-i}=\frac{n\omega-1}{n-1}=\frac{n}{n-1}\omega-\frac{1}{n-1}$ and Equation (3).

Since $r_1,\omega_Y,\omega_N\in(0,1)$, as $n\to\infty$,

$$\begin{aligned}
&P(\omega_{-i}=0\,|\,s_i=1)\\
=&P(\omega_{-i}=0\,|\,Y)P(Y\,|\,s_i=1)+P(\omega_{-i}=0\,|\,N)P(N\,|\,s_i=1)\\
=&r_1(1-\omega_Y)^{n-1}+(1-r_1)(1-\omega_N)^{n-1}\to 0,
\end{aligned}$$

and $P(\omega_{-i}=1\,|\,s_i=1)=r_1\omega_Y^{n-1}+(1-r_1)\omega_N^{n-1}\to 0$. We then have $\tilde{\omega}_1\to\bar{\omega}_1$. □

I.2. Proof of Lemma 2

Proof. By the weak law of large numbers, we have that for any $\delta>0$ and $\varepsilon>0$ there exists $n(\delta,\varepsilon)$ such that for all $n\geq n(\delta,\varepsilon)$:

$$P\left(\left|E\left[\omega_{-i}\,|\,Y\right]-\omega_{-i}\right|<\delta\,|\,Y\right)>1-\varepsilon. \tag{11}$$

Then note that $E[\omega_{-i} \,|\, Y] = \omega_Y$. For any $\varepsilon > 0$, taking $\delta = \omega_Y - \bar{\omega}_1 > 0$ in Equation (11), we have that for all $n \geq n_1 \equiv n(\omega_Y - \bar{\omega}_1, \varepsilon)$,

$$\begin{aligned} 1 - \varepsilon <& P\left(|\omega_Y - \omega_{-i}| < \omega_Y - \bar{\omega}_1 \,|\, Y\right) \\ \leq& P\left(\omega_Y - \omega_{-i} < \omega_Y - \bar{\omega}_1 \,|\, Y\right) = P(\omega_{-i} > \bar{\omega}_1 \,|\, Y). \end{aligned}$$

Similarly, for all $n \geq n_2 \equiv n(\bar{\omega}_0 - \omega_N, \varepsilon)$, $P(\omega_{-i} < \bar{\omega}_0 \,|\, N) > 1 - \varepsilon$.

From Lemma 1, for any $\delta > 0$, there exists $m(\delta)$ such that $|\tilde{\omega}_1 - \bar{\omega}_1| < \delta$ and $|\tilde{\omega}_0 - \bar{\omega}_0| < \delta$ for all $n \geq m(\delta)$. Let us first pick $\delta = \frac{\omega_Y - \bar{\omega}_1}{2} > 0$ and also in Equation (11). Then for all $n \geq n_3 \equiv \max\{n\left(\frac{\omega_Y - \bar{\omega}_1}{2}, \varepsilon\right), m(\frac{\omega_Y - \bar{\omega}_1}{2})\}$, $|\tilde{\omega}_1 - \bar{\omega}_1| < \frac{\omega_Y - \bar{\omega}_1}{2}$ and $P\left(|\omega_Y - \omega_{-i}| < \frac{\omega_Y - \bar{\omega}_1}{2} \,|\, Y\right) > 1 - \varepsilon$. Hence,

$$\begin{aligned} P(\omega_{-i} > \tilde{\omega}_1 \,|\, Y) &\geq P(\omega_{-i} > \frac{\omega_Y + \bar{\omega}_1}{2} > \tilde{\omega}_1 \,|\, Y) \\ &= P(\omega_{-i} > \frac{\omega_Y + \bar{\omega}_1}{2} \,|\, Y) \\ &= P(\omega_Y - \omega_{-i} < \frac{\omega_Y - \bar{\omega}_1}{2} \,|\, Y) \\ &\geq P\left(|\omega_Y - \omega_{-i}| < \frac{\omega_Y - \bar{\omega}_1}{2} \,|\, Y\right) > 1 - \varepsilon. \end{aligned}$$

Similarly, for all $n \geq n_4 \equiv \max\{n\left(\frac{\bar{\omega}_0 - \omega_N}{2}, \varepsilon\right), m(\frac{\bar{\omega}_0 - \omega_N}{2})\}$, $P(\omega_{-i} < \tilde{\omega}_0 \,|\, N) > 1 - \varepsilon$.

Finally, taking $n^* = \max\{n_1, n_2, n_3, n_4\}$ completes the proof. □

I.3. Proof of Lemma 3

Proof. We first note that $\tilde{\omega}_1 = \sum_{k=1}^{n-2} \frac{k}{n-1} \tilde{P}_1\left(\omega_{-i} = \frac{k}{n-1}\right)$ and

$$\begin{aligned} \tilde{E}_1\left[\omega_{-i} \,|\, \mathcal{E}_0^{-i}(p_i)\right] &= \sum_{k=1}^{n-2} \frac{k}{n-1} \tilde{P}_1\left(\omega_{-i} = \frac{k}{n-1} \,|\, \mathcal{E}_0^{-i}(p_i)\right) \\ &= \sum_{k=1}^{n-2} \frac{k}{n-1} \tilde{P}_1\left(\omega_{-i} = \frac{k}{n-1}\right) \frac{\tilde{P}_1\left(\mathcal{E}_0^{-i}(p_i) \,|\, \omega_{-i} = \frac{k}{n-1}\right)}{\tilde{P}_1\left(\mathcal{E}_0^{-i}(p_i)\right)} \quad (12) \end{aligned}$$

In what follows, we establish $\tilde{E}_1\left[\omega_{-i} \,|\, \mathcal{E}_0^{-i}(p_i)\right] \leq \tilde{\omega}_1$ by showing that $\tilde{P}_1(\omega_{-i})$ first-order stochastically dominates $\tilde{P}_1(\omega_{-i} \,|\, \mathcal{E}_0^{-i}(p_i))$.

Observe that

$$\begin{aligned}
&\tilde{P}_1\left(\mathcal{E}_0^{-i}\left(p_i\right)|\,\omega_{-i}=\frac{k}{n-1}\right)\\
=&\tilde{P}_1\left(\exists j\neq i,\ \text{s.t. } s_j=0 \text{ and } p_j\leq p_i\,|\,\omega_{-i}=\frac{k}{n-1}\right)\\
=&1-\tilde{P}_1\left(\forall j\neq i, s_j=0: p_j>p_i\,|\,\omega_{-i}=\frac{k}{n-1}\right)\\
=&1-\left(1-p_i\right)^{n-k-1}
\end{aligned}\tag{13}$$

is strictly decreasing in k. Hence, for the series $\{\tilde{P}_1\left(\mathcal{E}_0^{-i}\left(p_i\right)|\,\omega_{-i}=\frac{l}{n-1}\right)\}_{l=1}^{n-2}$, any weighted average of the first k terms is strictly larger than any weighted average of the last $n-k-2$ terms. We use this result to obtain that for integer $k\in[1,n-3]$,

$$\begin{aligned}
&\tilde{P}_1\left(\mathcal{E}_0^{-i}\left(p_i\right)|\,\omega_{-i}\leq\frac{k}{n-1}\right)\\
=&\sum_{l=1}^{k}\tilde{P}_1\left(\mathcal{E}_0^{-i}\left(p_i\right)|\,\omega_{-i}=\frac{l}{n-1}\right)\frac{\tilde{P}_1\left(\omega_{-i}=\frac{l}{n-1}\right)}{\tilde{P}_1\left(\omega_{-i}\leq\frac{k}{n-1}\right)}\\
=&\tilde{P}_1\left(\omega_{-i}\leq\frac{k}{n-1}\right)\sum_{l=1}^{k}\tilde{P}_1\left(\mathcal{E}_0^{-i}\left(p_i\right)|\,\omega_{-i}=\frac{l}{n-1}\right)\frac{\tilde{P}_1\left(\omega_{-i}=\frac{l}{n-1}\right)}{\tilde{P}_1\left(\omega_{-i}\leq\frac{k}{n-1}\right)}\\
&+\tilde{P}_1\left(\omega_{-i}>\frac{k}{n-1}\right)\sum_{l=1}^{k}\tilde{P}_1\left(\mathcal{E}_0^{-i}\left(p_i\right)|\,\omega_{-i}=\frac{l}{n-1}\right)\frac{\tilde{P}_1\left(\omega_{-i}=\frac{l}{n-1}\right)}{\tilde{P}_1\left(\omega_{-i}\leq\frac{k}{n-1}\right)}\\
>&\tilde{P}_1\left(\omega_{-i}\leq\frac{k}{n-1}\right)\sum_{l=1}^{k}\tilde{P}_1\left(\mathcal{E}_0^{-i}\left(p_i\right)|\,\omega_{-i}=\frac{l}{n-1}\right)\frac{\tilde{P}_1\left(\omega_{-i}=\frac{l}{n-1}\right)}{\tilde{P}_1\left(\omega_{-i}\leq\frac{k}{n-1}\right)}\\
&+\tilde{P}_1\left(\omega_{-i}>\frac{k}{n-1}\right)\sum_{l=k+1}^{n-2}\tilde{P}_1\left(\mathcal{E}_0^{-i}\left(p_i\right)|\,\omega_{-i}=\frac{l}{n-1}\right)\frac{\tilde{P}_1\left(\omega_{-i}=\frac{l}{n-1}\right)}{\tilde{P}_1\left(\omega_{-i}>\frac{k}{n-1}\right)}\\
=&\sum_{l=1}^{n-2}\tilde{P}_1\left(\mathcal{E}_0^{-i}\left(p_i\right)|\,\omega_{-i}=\frac{l}{n-1}\right)\tilde{P}_1\left(\omega_{-i}=\frac{l}{n-1}\right)\\
=&\tilde{P}_1\left(\mathcal{E}_0^{-i}\left(p_i\right)\right),
\end{aligned}\tag{14}$$

where the inequality is obtained because $\frac{\tilde{P}_1\left(\omega_{-i}=\frac{l}{n-1}\right)}{\tilde{P}_1\left(\omega_{-i}\leq\frac{k}{n-1}\right)}=\frac{\tilde{P}_1\left(\omega_{-i}=\frac{l}{n-1}\right)}{\sum_{l=1}^{k}\tilde{P}_1\left(\omega_{-i}=\frac{l}{n-1}\right)}$ and thus the expression $\sum_{l=1}^{k}\tilde{P}_1\left(\mathcal{E}_0^{-i}\left(p_i\right)|\,\omega_{-i}=\frac{l}{n-1}\right)\frac{\tilde{P}_1\left(\omega_{-i}=\frac{l}{n-1}\right)}{\tilde{P}_1\left(\omega_{-i}\leq\frac{k}{n-1}\right)}$ is a weighted average of the first k terms of $\{\tilde{P}_1\left(\mathcal{E}_0^{-i}\left(p_i\right)|\,\omega_{-i}=\frac{l}{n-1}\right)\}_{l=1}^{n-2}$. When $k=n-2$, $\tilde{P}_1\left(\mathcal{E}_0^{-i}\left(p_i\right)|\,\omega_{-i}\leq\frac{k}{n-1}\right)=$

$\tilde{P}_1\left(\mathcal{E}_0^{-i}(p_i)\right)$. These imply that for any $k \in [1, n-3]$,

$$\begin{aligned}\tilde{P}_1\left(\omega_{-i} \leq \frac{k}{n-1} \,|\, \mathcal{E}_0^{-i}(p_i)\right) =& \tilde{P}_1\left(\omega_{-i} \leq \frac{k}{n-1}\right) \frac{\tilde{P}_1\left(\mathcal{E}_0^{-i}(p_i) \,|\, \omega_{-i} \leq \frac{k}{n-1}\right)}{\tilde{P}_1\left(\mathcal{E}_0^{-i}(p_i)\right)} \\ >& \tilde{P}_1\left(\omega_{-i} \leq \frac{k}{n-1}\right).\end{aligned}$$

Hence, by first order stochastic dominance of the unconditional distribution of ω_{-i} over the distribution conditioning on $\mathcal{E}_0^{-i}$, we get $\tilde{E}_1\left[\omega_{-i} \,|\, \mathcal{E}_0^{-i}(p_i)\right] < \tilde{\omega}_1$.

$\tilde{E}_1\left[\omega_{-i} \,|\, \mathcal{E}_1^{-i}(p_i)\right] > \tilde{\omega}_1$ is similarly established by showing that the unconditional distribution ω_{-i} is stochastically dominated by the distribution conditioning on $\mathcal{E}_1^{-i}$. This is obtained from

$$\begin{aligned}&\tilde{P}_1\left(\mathcal{E}_1^{-i}(p_i) \,|\, \omega_{-i} = \frac{k}{n-1}\right) \\ =&\tilde{P}_1\left(\exists j \neq i, \text{ s.t. } s_j = 1 \text{ and } p_j \geq p_i \,|\, \omega_{-i} = \frac{k}{n-1}\right) \\ =&1 - \tilde{P}_1\left(\forall j \neq i,\, s_j = 1 : p_j < p_i \,|\, \omega_{-i} = \frac{k}{n-1}\right) \\ =&1 - p_i^k\end{aligned} \tag{15}$$

being strictly increasing in k. □

I.4. Proof of Lemma 4

Proof. Consider Equation (12). Only the weights $\frac{\tilde{P}_1\left(\mathcal{E}_0^{-i}(p_i) \,|\, \omega_{-i} = \frac{k}{n-1}\right)}{\tilde{P}_1\left(\mathcal{E}_0^{-i}(p_i)\right)}$ are functions of p_i. Since the denominator is the weighted sum of the numerator for all k, it is enough to show that $\tilde{P}_1\left(\mathcal{E}_0^{-i}(p_i) \,|\, \omega_{-i} = \frac{k}{n-1}\right)$ is continuous in p_i. This is immediate from Equation (13). □

I.5. Proof of Lemma 5

Proof. We start from Equation (12) and first show that the denominator of the fraction in the expression goes to 1 as $n \to \infty$. Using Equation (13) and since for any integer $k \in [1, n-2]$,

$$\tilde{P}_1\left(\omega_{-i} = \frac{k}{n-1}\right) \\ = \binom{n-1}{k}\omega_Y^k\left(1-\omega_Y\right)^{n-k-1}\tilde{r}_1 + \binom{n-1}{k}\omega_N^k\left(1-\omega_N\right)^{n-k-1}\left(1-\tilde{r}_1\right), \tag{16}$$

we arrive at

$$\begin{aligned}
\tilde{P}_1\left(\mathcal{E}_0^{-i}\left(p_i\right)\right) =& \sum_{k=1}^{n-2} \tilde{P}_1\left(\mathcal{E}_0^{-i}\left(p_i\right) \mid \omega_{-i} = \frac{k}{n-1}\right)\tilde{P}_1\left(\omega_{-i} = \frac{k}{n-1}\right) \\
=& \sum_{k=1}^{n-2}\left[1-\left(1-p_i\right)^{n-k-1}\right]\tilde{P}_1\left(\omega_{-i} = \frac{k}{n-1}\right) \\
=& 1 - \sum_{k=1}^{n-2}\left(1-p_i\right)^{n-k-1}\binom{n-1}{k}\left[\omega_Y^k\left(1-\omega_Y\right)^{n-k-1}\ \tilde{r}_1 + \omega_N^k\left(1-\omega_N\right)^{n-k-1}\ \left(1-\tilde{r}_1\right)\right] \\
=& 1 - \tilde{r}_1\{\left[\omega_Y + \left(1-p_i\right)\left(1-\omega_Y\right)\right]^{n-1} - \omega_Y^{n-1} - (1-p_i)^{n-1}(1-\omega_Y)^{n-1}\} \\
& - \left(1-\tilde{r}_1\right)\{\left[\omega_N + \left(1-p_i\right)\left(1-\omega_N\right)\right]^{n-1} - \omega_N^{n-1} - (1-p_i)^{n-1}(1-\omega_N)^{n-1}\},
\end{aligned} \tag{17}$$

where the last equality is obtained from the binomial theorem. In the limit, for any $p_i \in (0,1)$,

$$\lim_{n\to\infty} \tilde{P}_1\left(\mathcal{E}_0^{-i}\left(p_i\right)\right) = 1 \tag{18}$$

since p_i, ω_Y, and ω_N are all strictly between 0 and 1.

Next, we bound the expression in Equation (12) from above and below to show that it tends to $\tilde{\omega}_1$ in the limit. Lemma 3 already gives the upper bound, $\tilde{\omega}_1$. For the lower bound:

$$\begin{aligned}
& \tilde{E}_1\left[\omega_{-i} \,|\, \mathcal{E}_0^{-i}\left(p_i\right)\right] \\
=& 1 - \tilde{E}_1\left[1-\omega_{-i} \,|\, \mathcal{E}_0^{-i}\left(p_i\right)\right] \\
=& 1 - \frac{\sum_{k=1}^{n-2}\left(1-\frac{k}{n-1}\right)\tilde{P}_1\left(\mathcal{E}_0^{-i}\left(p_i\right) \mid \omega_{-i} = \frac{k}{n-1}\right)\tilde{P}_1\left(\omega_{-i} = \frac{k}{n-1}\right)}{\tilde{P}_1\left(\mathcal{E}_0^{-i}\left(p_i\right)\right)} \\
>& 1 - \frac{\sum_{k=1}^{n-2}\left(1-\frac{k}{n-1}\right)\tilde{P}_1\left(\omega_{-i} = \frac{k}{n-1}\right)}{\tilde{P}_1\left(\mathcal{E}_0^{-i}\left(p_i\right)\right)} \\
=& 1 - \frac{1-\tilde{\omega}_1}{\tilde{P}_1\left(\mathcal{E}_0^{-i}\left(p_i\right)\right)}.
\end{aligned} \tag{19}$$

Combining the two bounds, we get

$$1-\frac{1-\tilde{\omega}_1}{\tilde{P}_1\left(\mathcal{E}_0^{-i}\left(p_i\right)\right)}<\tilde{E}_1\left[\omega_{-i}\,|\,\mathcal{E}_0^{-i}\left(p_i\right)\right]<\tilde{\omega}_1. \tag{20}$$

Taking the limit and using Equation (18) and Lemma 1, we get that for any $p_i \in (0,1)$, $\lim_{n\to\infty}\tilde{E}_1\left[\omega_{-i}\,|\,\mathcal{E}_0^{-i}\left(p_i\right)\right]=\lim_{n\to\infty}\tilde{\omega}_1=\bar{\omega}_1$. □

Online Appendix II: Simulations on Proposition 3

The simulation procedure for Proposition 3 is as follows: Given a group size $n \in \{20, 40, 60, 80, 100, 200, 300\}$ and signal technology parameters $r = \frac{1}{2}$ and any (ω_N, ω_Y) pair[18] randomly selected from $\{0.05, 0.1, ..., 0.95\}$, we can calculate $\tilde{E}_1\left[\omega_{-i} \,|\, \mathcal{E}_0^{-i}(p_i)\right]$ and $\tilde{E}_0\left[\omega_{-i} \,|\, \mathcal{E}_1^{-i}(p_i)\right]$ as a function of p_i.[19] This means that we know the trading

[18]This results in $\binom{19}{2} = 171$ pairs of ω_Y and ω_N where $\omega_Y > \omega_N$.

[19]For signal-1 agents, we first start from Equation (12):

$$
\begin{aligned}
\tilde{E}_1\left[\omega_{-i} \,|\, \mathcal{E}_0^{-i}(p_i)\right] =& \sum_{k=1}^{n-2} \frac{k}{n-1} \tilde{P}_1\left(\omega_{-i} = \frac{k}{n-1}\right) \frac{\tilde{P}_1\left(\mathcal{E}_0^{-i}(p_i) \,|\, \omega_{-i} = \frac{k}{n-1}\right)}{\tilde{P}_1\left(\mathcal{E}_0^{-i}(p_i)\right)} \\
=& \frac{1}{n-1} \frac{1}{\tilde{P}_1\left(\mathcal{E}_0^{-i}(p_i)\right)} \sum_{k=1}^{n-2} k \tilde{P}_1\left(\omega_{-i} = \frac{k}{n-1}\right) \tilde{P}_1\left(\mathcal{E}_0^{-i}(p_i) \,|\, \omega_{-i} = \frac{k}{n-1}\right),
\end{aligned}
$$

where $\tilde{P}_1\left(\omega_{-i} = \frac{k}{n-1}\right)$ is defined in Equation (16) as

$$
\tilde{P}_1\left(\omega_{-i} = \frac{k}{n-1}\right) = \binom{n-1}{k} \omega_Y^k (1-\omega_Y)^{n-k-1} \tilde{r}_1 + \binom{n-1}{k} \omega_N^k (1-\omega_N)^{n-k-1} (1-\tilde{r}_1),
$$

$\tilde{P}_1\left(\mathcal{E}_0^{-i}(p_i) \,|\, \omega_{-i} = \frac{k}{n-1}\right)$ is defined in Equation (13) as

$$
\tilde{P}_1\left(\mathcal{E}_0^{-i}(p_i) \,|\, \omega_{-i} = \frac{k}{n-1}\right) = 1 - (1-p_i)^{n-k-1},
$$

and $\tilde{P}_1\left(\mathcal{E}_0^{-i}(p_i)\right)$ is defined in Equation (17) as

$$
\begin{aligned}
\tilde{P}_1\left(\mathcal{E}_0^{-i}(p_i)\right) =& 1 - \tilde{r}_1\{[\omega_Y + (1-p_i)(1-\omega_Y)]^{n-1} - \omega_Y^{n-1} - (1-p_i)^{n-1}(1-\omega_Y)^{n-1}\} \\
& - (1-\tilde{r}_1)\{[\omega_N + (1-p_i)(1-\omega_N)]^{n-1} - \omega_N^{n-1} - (1-p_i)^{n-1}(1-\omega_N)^{n-1}\}.
\end{aligned}
$$

Then we derive the expression of $\tilde{r}_1$ as a function of the signal technology parameters:

$$
\begin{aligned}
\tilde{r}_1 =& \tilde{P}_1(Y) \\
=& P\left(Y \,|\, s_i = 1, \omega_{-i} \in (0,1)\right) \\
=& \frac{P\left(\omega_{-i} \in (0,1) \,|\, Y\right) P\left(Y \,|\, s_i = 1\right)}{P\left(\omega_{-i} \in (0,1) \,|\, Y\right) P\left(Y \,|\, s_i = 1\right) + P\left(\omega_{-i} \in (0,1) \,|\, N\right) P\left(N \,|\, s_i = 1\right)} \\
=& \frac{r_1\left(1 - (1-\omega_Y)^{n-1} - \omega_Y^{n-1}\right)}{r_1\left(1 - (1-\omega_Y)^{n-1} - \omega_Y^{n-1}\right) + (1-r_1)\left(1 - (1-\omega_N)^{n-1} - \omega_N^{n-1}\right)},
\end{aligned}
$$

where r_1 is defined in Equation (1) as

$$
r_1 = \frac{r\omega_Y}{r\omega_Y + (1-r)\omega_N}.
$$

strategies of both buyers and sellers. We then run 200 simulations for a group size n for both states Y and N. During each simulation under state S, we first determine the signal each agent gets using the sampling probability ω_S and we assume that agents endorses truthfully ($e_i = s_i$). Then we randomly draw a price p_i from a uniform distribution on $(0, 1)$ for each agent. Comparing their expectations of the asset value with the price, we determine whether each agent is willing to bet or not (d_i being 1 or 0). Then we check the trading conditions and determine whether the trade goes through. Payoffs are determined and we document if agents endorsing the actual states of nature, and only them, make a profit. We then calculate the proportion of simulations that satisfy this property.

To see whether our FTM-A algorithm could detect the actual state of nature in each simulation, we only take the payoffs of all the agents and compare the average payoffs of the two sides. In case of no trade or equal average payoffs, we treat the performance of FTM-A as 0.5, due to the prior r being 0.5. For our FTM-L algorithm, we take only the data of agents' endorsements e_i, prices p_i, and betting decisions d_i, and use these to fit the logistic curves.

Combining all of these equations, we can express $\tilde{E}_1\left[\omega_{-i}\,|\,\mathcal{E}_0^{-i}(p_i)\right]$ explicitly as a function of p_i and signal technology parameters r, ω_Y, and ω_N.

We used MATLAB to calculate the conditional expectations. Operationally, the MATLAB function "binopdf", required to calculate $\binom{n-1}{k}\omega_S^k\omega_S^{n-k-1}$, is too slow for the large number of simulations. Hence, we first use the inequalities in Equation (20) to compute the upper and lower bounds of $\tilde{E}_1\left[\omega_{-i}\,|\,\mathcal{E}_0^{-i}(p_i)\right]$: $1-\frac{1-\tilde{\omega}_1}{\tilde{P}_1\left(\mathcal{E}_0^{-i}(p_i)\right)} < \tilde{E}_1\left[\omega_{-i}\,|\,\mathcal{E}_0^{-i}(p_i)\right] < \tilde{\omega}_1$, where

$$\tilde{\omega}_1 = \sum_{k=1}^{n-2}\frac{k}{n-1}\tilde{P}_1\left(\omega_{-i}=\frac{k}{n-1}\right) = \tilde{r}_1\left(\omega_Y-\omega_Y^{n-1}\right)+\left(1-\tilde{r}_1\right)\left(\omega_N-\omega_N^{n-1}\right),$$

and we used the expression of the expected value of a binomial distribution: $\sum_{k=0}^{n-1}k\binom{n-1}{k}\omega_S^k\left(1-\omega_S\right)^{n-k-1} = (n-1)\omega_S$. Only when p_i is within these bounds do we then calculate the exact values of the expectation to determine the trading decision.

Similarly, for signal-0 agents, we can also express $\tilde{E}_0\left[\omega_{-i}\,|\,\mathcal{E}_1^{-i}(p_i)\right]$ explicitly as a function of p_i, group size n, and signal technology parameters r, ω_Y, and ω_N, by following similar steps. Operationally speaking, we can replace s_i, r, ω_Y, ω_N, ω_{-i}, p_i by $1-s_i$, $1-r$, $1-\omega_N$, $1-\omega_Y$, $1-\omega_{-i}$, $1-p_i$, respectively, and calculate $\tilde{E}_1\left[\omega_{-i}\,|\,\mathcal{E}_0^{-i}(p_i)\right]$ instead.

Online Appendix III: Experimental details

III.1. Parameter values for each task

In our current experimental design, $P(S = Y) = P(S = N) = \frac{1}{2}$, and group size $n = 100$. Table 3 lists parameter values ω_N and ω_Y of the 30 tasks in our experiment. They contain all combinations where $(\omega_N, \omega_Y) \in \{0.05, 0.1, ..., 0.45, 0.55, ..., 0.95\}$, $\omega_Y - \omega_N \in \{0.2, 0.25, 0.3\}$, and $\bar{\omega}_1 - \bar{\omega}_0 > 0.04$.

In contrast to the finite case in our theoretical setting (Section 2.4), the settlement value does not exclude the participant's own endorsement, and the number of participants receiving signal 1 is fixed at $n\omega_S$ for $S = \{Y, N\}$. Therefore, there is no uncertainty stemming from sampling and hence under the truth-telling equilibrium,

$$\begin{aligned}
&E\left[v_i \,|\, s_i = 1, v_i \in (0,1), \mathcal{E}_0^{-i}(p_i)\right] \\
=&\,\omega_Y P(v_i = \omega_Y \,|\, s_i = 1, \mathcal{E}_0^{-i}(p_i)) + \omega_N P(v_i = \omega_N \,|\, s_i = 1, \mathcal{E}_0^{-i}(p_i)) \\
=&\,\omega_N + (\omega_Y - \omega_N) P(Y \,|\, s_i = 1, \mathcal{E}_0^{-i}(p_i)).
\end{aligned} \tag{21}$$

With this simplified expression, we can write the expectation as an explicit function of the price p_i, given $n = 100$ and parameters values ω_Y and ω_N for each task. First, using Bayes' rule, we get:

$$\begin{aligned}
&P(Y \,|\, s_i = 1, \mathcal{E}_0^{-i}(p_i)) \\
=&\frac{\omega_Y\left(1 - (1-p_i)^{n(1-\omega_Y)}\right)}{\omega_Y\left(1 - (1-p_i)^{n(1-\omega_Y)}\right) + \omega_N\left(1 - (1-p_i)^{n(1-\omega_N)}\right)},
\end{aligned} \tag{22}$$

where we used that $P(S = Y) = P(S = N) = \frac{1}{2}$, $P(s_i = 1 \,|\, S) = \omega_S$ for $S = \{Y, N\}$. Plugging this expression in Equation (21), we define, for each task, a reservation price p_1^* for buyers, such that $E\left[v_i \,|\, s_i = 1, v_i \in (0,1), \mathcal{E}_0^{-i}(p_i)\right] > p_i$ when $p_i < p_1^*$, $E\left[v_i \,|\, s_i = 1, v_i \in (0,1), \mathcal{E}_0^{-i}(p_1^*)\right] = p_1^*$, and $E\left[v_i \,|\, s_i = 1, v_i \in (0,1), \mathcal{E}_0^{-i}(p_i)\right] < p_i$ when $p_i > p_1^*$. Similarly, we can also define a reservation price p_0^* for sellers.

If p_1^* and p_0^* exist, and are such that $p_1^* > p_0^*$, then signal-1 agents will buy at price $p \leq p_1^*$ and signal-0 agents will sell at price $p \geq p_0^*$. These strategies constitute an equilibrium.

Table 3: Task parameter values

set	task	ω_N	ω_Y	p_0^*	p_1^*
1	1	0.05	0.25	0.14	0.22
2	2	0.05	0.3	0.16	0.26
1	3	0.05	0.35	0.17	0.31
2	4	0.1	0.3	0.19	0.25
1	5	0.1	0.35	0.2	0.29
2	6	0.1	0.4	0.22	0.34
1	7	0.15	0.35	0.24	0.29
2	8	0.15	0.4	0.25	0.33
1	9	0.15	0.45	0.27	0.37
2	10	0.2	0.4	0.29	0.33
1	11	0.2	0.45	0.3	0.37
2	12	0.25	0.45	0.33	0.38
2	13	0.75	0.95	0.78	0.86
1	14	0.7	0.95	0.74	0.84
2	15	0.65	0.95	0.69	0.83
1	16	0.7	0.9	0.75	0.81
2	17	0.65	0.9	0.71	0.8
1	18	0.6	0.9	0.66	0.78
2	19	0.65	0.85	0.71	0.76
1	20	0.6	0.85	0.67	0.75
2	21	0.55	0.85	0.63	0.73
1	22	0.6	0.8	0.67	0.71
2	23	0.55	0.8	0.63	0.7
1	24	0.55	0.75	0.62	0.67
1	25	0.3	0.6	0.41	0.5
2	26	0.35	0.6	0.45	0.51
1	27	0.35	0.65	0.46	0.54
2	28	0.4	0.6	0.48	0.52
1	29	0.4	0.65	0.49	0.55
2	30	0.4	0.7	0.5	0.59

We derive p_1^* and p_0^* for all the 30 tasks, which are shown in the last two columns of Table 3. Note that $\overline{\omega}_1$ and $\overline{\omega}_0$ defined in Equations (3)-(4) yields essentially the same values up to the second decimal place, except tasks 9 and 27 for buyers, but still $|p_1^* - \overline{\omega}_1| < 0.01$. Thus, conditioning on the occurrence of trade is negligible and for the parameters chosen in the experiment, the incentives of the Bayesian market are essentially as explained in Section 2.1 where n is infinite.

III.2. Truth-telling at the individual level

Table 4 shows the proportion of participants with at least certain numbers of truth-telling in both SPA and FTM treatments.

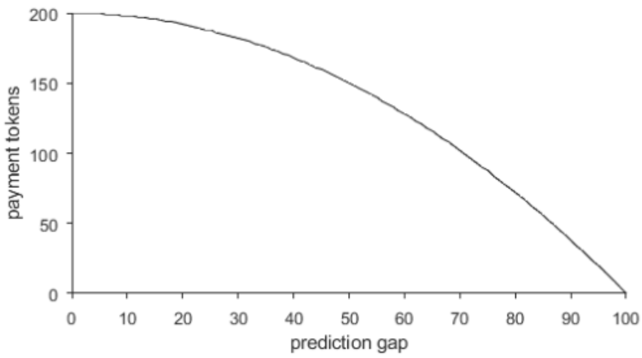
Table 4: Proportion of participants with at least certain numbers of truth-telling

at least	SPA	FTM
1	99.8%	100.0%
2	99.8%	99.8%
3	99.8%	99.8%
4	99.6%	99.6%
5	98.9%	98.9%
6	98.1%	98.3%
7	96.1%	96.6%
8	94.6%	94.1%
9	91.3%	91.3%
10	88.7%	85.8%
11	84.0%	82.2%
12	79.0%	77.6%
13	72.7%	71.9%
14	63.4%	64.3%
15	50.6%	53.1%

III.3. Experimental instructions and training rounds

All participants first watched the experimental instruction video (YouTube link) where the experimental setting of urns and balls was explained. Then they went through five rounds of training and got feedback about how the payment was calculated. Figure 8 shows an example from the SPA treatment, and Figure 9 shows an example from the FTM treatment where the bet went through.

Your payment for this task depends on your prediction gap, which is the distance between your prediction and the actual number of participants who have guessed the same urn as you. **The less the prediction gap is, the higher your payment will be.**



Your payment for this task: 182 tokens

At the end of the experiment, we sum up your total payment tokens from all the 15 tasks and determine your bonus. Your bonus (in £) equals the total number of payment tokens divided by 1000. You could earn up to £3 in bonus.

Figure 8: An example of feedback in the training rounds in SPA.

The earnings of the bet (in tokens) is this number minus 55:
Earnings = 80 - 55 = 25

To make sure you don't lose money (if the earnings are negative), you will be endowed with 100 tokens, whether you take the bet or not.

You chose to take the bet.

Your bet goes through if someone took the opposite bet, which means, someone bet that less than 55 participants chose Urn Right (or in other words, more than 45 chose the other urn).

Your bet went through. Your payment for this task is 100 + earnings = 125 tokens.

If you had not taken the bet, then your payment would have been the endowment, 100 tokens.

At the end of the experiment, we sum up your total payment tokens from all the 15 tasks and determine your bonus. Your bonus (in £) equals the total number of payment tokens divided by 1000. You could earn up to £3 in bonus.

Note that the value 55 in the bet is randomly drawn.

Figure 9: An example of feedback in the training rounds in FTM.