Aggregation and Manipulation in Prediction Markets: Effects of Trading Mechanism and Information Distribution

[Working Paper][*]

Lian Jian
School of Information
University of Michigan
Ann Arbor, MI 48109, USA
ljian@umich.edu

Rahul Sami
School of Information
University of Michigan
Ann Arbor, MI 48109, USA
rsami@umich.edu

ABSTRACT

We conduct laboratory experiments on variants of market scoring rule prediction markets, under different information distribution patterns, in order to evaluate the efficiency and speed of information aggregation, as well as test recent theoretical results on manipulative behavior by traders. We find that markets structured to have a fixed sequence of trades exhibit greater accuracy of information aggregation than the typical form that has unstructured trade. Prior theoretical predictions of differing strategic behavior under complementary information distributions and substitute information distributions are confirmed when the trading order is structured, but not in markets with an unstructured trading order. In the case of the markets with a structured order, we find that the information aggregation is consequently slower when information is complementary, as traders more frequently engage in bluffing and delaying strategies. In comparing two commonly used mechanisms, we find no significant difference between the performance of the direct probability-report form and the indirect security-trading form of the market scoring rule.

1. INTRODUCTION

Market prices facilitate efficient resource allocations, and also act as information aggregators: they reflect market participants' valuation of resources [15]. Prediction markets — markets in which traders buy and sell bets on future events — are designed to explicitly take advantage of the information aggregation function of market prices to provide decision makers with forecasts of future events. Such markets have been created for a wide range of applications; examples include the Iowa Electronic Market for forecasting elections and other political events, the Hollywood Stock Exchange for forecasting movie box office receipts, and intra-company markets to forecast sales. In this paper, we use laboratory experiments to study the effectiveness of information aggregation of different variants of market scoring rule prediction markets, under differing information conditions. Our experimental results shed new light on the validity of theoretical predictions for these markets, as well as on the impact of common mechanism variations.

Although all prediction markets involve speculative bets on future events, the particular form taken by these bets can vary significantly. One common form is the *continuous double auction*, in which traders submit buy or sell orders for units of a security, and the market operator matches buy and sell orders to execute trades. When the outcome of the future event is known, the security is cashed out at a value that depends on the outcome. Continuous double auctions are very complex strategically, for the traders as well as the analyst. Recently, a new market form for prediction markets, the market scoring rule [13], has become popular. Market scoring rules (MSR) are being used in a growing number of deployed prediction markets, including prominently the public prediction market site *Inkling Markets* (inklingmarkets.com). MSR markets have advantages over continuous double auction markets, particularly in situations with thin trade. In addition, they are more amenable to theoretical analysis, and there have been a number of recent studies that provide insight into optimal theoretical strategies in MSR markets [13, 5, 8, 3]. In this paper, we use human-subject laboratory experiments to study the speed and efficiency of information aggregation in MSR markets, while varying the mechanism form, constraints on trade timing, and information distribution pattern.

The first dimension of variation we consider is in comparing two commonly used mechanism forms that implement MSR markets: a direct mechanism in which traders report their beliefs as probabilities, and an indirect mechanism in which traders reveal their beliefs through buying and selling securities. The simplest representation of a market scoring rule market is as a sequence of *reported probabilities.* Each trade in the market involves a trader changing the current report. We call this type of market a *direct MSR* market. Once the outcome of the event is revealed, each trader is paid off according to a prespecified scoring rule, which depends on his report as well as the previous report. In practice, however, markets that use the market scoring rule, such as the public prediction market site Inklingmarkets.com, typically use an alternative mechanism: Traders buy and sell units of a security, but instead of trading directly with each other, they trade with an *automated market maker* who constantly adjusts the prices. We call this type of market an *indirect*

[*]This paper will be presented at the 2010 ACM Conference on Electronic Commerce (EC'10), June 7–11, 2010, Cambridge, Massachusetts, USA. This work was supported by the National Science Foundation grant CCF-0728768. We thank our research assistants Matthew Bombyk and Roberto Jimenez.

MSR market. Direct and indirect MSR markets are formally equivalent, but may appear very different to traders in the market. There is an active debate about which interface is more effective in practice [20]; our laboratory experiments provide insight into this question.

Market speed and efficiency depends on appropriate behavior by traders, and hence trader strategies have to be taken into account. Hanson [13] showed that MSR markets have a myopic honesty property: A trader trading only once maximizes her expected profit by reporting her true beliefs. However, for traders potentially using non-myopic strategies over multiple trades, the theoretical results show a sharp distinction based on the pattern of information distribution among the traders: If traders' information signals are independent *conditional on the true outcome of the event*, then the signals are substitutes, and honest reporting of beliefs is the optimal strategy even in a non-myopic sense; if the traders' information signals are *unconditionally independent*, then the signals are complements, and honest reporting is in general not a sequential equilibrium [3]. In the latter case, a complete characterization of equilibria is unknown, but it is known that a trader can profitably deviate from the honest strategy profile by bluffing (trading in the opposite direction to her signal with some probability) or delaying (waiting for the other traders to reveal their information before trading); these deviations are construed as manipulative strategies. This motivates the second dimension along which we vary our experimental design: We study market performance under a complementary signal structure, and under a substitute signal structure. This enables us to paint a broader picture of the comparison between different market forms, as well as to conduct the first experimental test of these theoretical results on strategic manipulation.

The third variation we study is in providing the structure of strictly sequenced opportunities to trade, as compared to the standard approach of letting traders choose when to trade in an unstructured way. We have two motivations in considering a structure: First, the existing theoretical results [5, 8, 3] implicitly assume a structured order of trading opportunities, and this experiment allows us to test if this assumption is of practical significance. Second, enforcing more structured interaction has been shown to help in group forecasting performance [12]. Our experiments allow us to test if the additional structure of a trading sequence, which might simplify traders' information processing, improves the aggregation performance of a prediction market.

We designed and carried out market trading experiments for each of the 8 treatments generated by a factorial exploration of these three dimensions of variation. All experiments reported here involved markets with two traders. This simplifies the signal interpretation problem for the subjects, thereby giving us a best-case situation in which to test the theoretical predictions. Based on the theoretical results summarized in [3], we expect the following: In the substitutes markets, traders should trade honestly, as early as they can. In the complements markets, traders have an incentive to reveal their information as late as possible; they also have an incentive to bluff with some probability if they can correct the market later. Both these should lead to poorer early information aggregation in the complements case than in the substitutes case. Further, with ideal rational traders, the choice of direct or indirect mechanism should make no difference to the market aggregation. The comparison of structured and unstructured trading orders is an open question.

The results of our experiments make several contributions to our understanding of the aggregative and strategic properties of prediction markets with different trading mechanisms, and under different information conditions. First, we find that structured markets (with an exogenous sequence of trading opportunities), aggregate information more efficiently than unstructured markets, with an endogenous trading order. (This result was significant in three out of the four treatments. For the fourth treatment, the comparison was in the same direction, but not statistically significant).

Second, in the first experimental comparison between the direct and indirect trading mechanisms that have been proposed for market scoring rules, we find no significant difference in performance. Third, in testing the theoretical results on the effect of information distribution on manipulative strategies, we find that they are borne out for the structured market with exogenous sequence of trading opportunities (this result is significant for the direct structured market, and of borderline significance for the indirect structured market), but not for the unstructured markets. This suggests that the timing and ordering of trades is an important feature to include in future theoretical research in this area.

The rest of this paper is structured as follows: In section 2, we summarize the prior research related to our work. Section 3 details our experimental design, analysis metrics, and hypotheses. We present the results of our experiments in section 4. We summarize the paper and outline important directions for future work in section 5.

2. RELATED WORK AND BACKGROUND

The theoretical underpinnings of using market prices as reliable forecasts of future events are provided by the theory of Rational Expectations Equilibrium [18, 23, 9]. Rational expectations equilibrium models predict that, generically, prices in prediction markets can fully aggregate all individual traders' private information. Prediction markets' advantages over other methods of information aggregation such as polls and expert deliberations have also been empirically demonstrated in a large number of markets [2, 7, 26, 11]. Because of their perceived accuracy, as well as the fact that they are relatively easy and inexpensive to run, we are witnessing a rapid growth in the use of prediction markets as tools for information aggregation [7, Footnote 2].

2.1 Market Scoring Rules

In our study, we focus on market scoring rules (MSR) based prediction markets as suggested by [13]. Hanson outlined two alternative implementations of the MSR [13]. One is a direct implementation of the MSR (we call it direct MSR), in which each trader reports their own predictions and receives payments accordingly. The other one is an indirect implementation of the MSR (we call it indirect MSR), which contains a market maker offering n securities each of which pays \$1 if the associated outcome is realized [13, 4] and \$0 otherwise. The two implementations are mathematically and hence strategically equivalent, but they have very different look and feel to the market traders. Although both implementations have been used in practice [20], to the best of our knowledge, there have been no empirical tests comparing the performance of these two implementations to provide guidelines for prediction market designers. It is one of our goals in this paper to compare the performance of these two implementations.

Direct MSR. Scoring rules are tools for eliciting private beliefs. Given a random variable X which has n possible outcomes, to elicit an individual's, say Alice's, belief about the probabilities of each of these outcomes $p = (p_1, \cdots, p_n)$, we can ask her to express her beliefs by $r = (r_1, \cdots, r_n)$ — a vector of reported probabilities for the random variable X — and pay her based on the scoring rule $S = \{s_1(r), \cdots, s_n(r)\}$. Thus, if outcome 1 is realized, she will be paid $s_1(r)$; if 2 is realized, she will be paid $s_2(r)$, and so on. Alice maximizes her expected score $S(r)$ by choosing an r to report:

$$S(r) = \Sigma_{i=1}^{n} s_i(r) p_i \tag{1}$$

If the scoring rule is *proper*, Alice would find that $r = p$ maximizes her expected payoffs expressed in Equation (1). Popularly used proper scoring rules include quadratic, spherical, and logarithmic scoring rules.[1] In an MSR-based prediction market, traders report their forecasts sequentially, and have access to the sequence of forecasts made up to the current time. A trader earns the difference between her score and the previous trader's score. That is, if outcome i is realized, trader m who reported r_m will receive payment $s_i(r_m) - s_i(r_{m-1})$, where r_{m-1} is the report of the previous trader. Throughout this paper, we use an MSR based on the **logarithmic** market scoring rule.

Indirect MSR. The market scoring rule can be viewed as a specific form of *automated market-maker*, an agent that posts prices, is always willing to trade securities at the posted price, and updates the prices following every trade. Thus, every trade in an MSR market is made with the automated market-maker as either the buyer or the seller; this is formally equivalent to the sequence model of direct MSR, while providing users with a mechanism that is more familiar to them from other markets. In particular, the (logarithmic) MSR equivalent market price of security i, p_i, can be derived from the following expression [1]:

$$p_i = \frac{e^{s_i/b}}{\sum_k e^{s_k/b}} \tag{2}$$

where b is the scaling factor in the scoring rule, and s_i is the total amount of security i that has been sold. Berg *et al.* [1] detail the implementation of indirect MSR markets. (For reviewers' reference, the direct and indirect MSR mechanisms used in our experiment have been included in Figure 2 and Figure 3 in Section A in the Appendix.)

2.2 Theoretical Analysis of MSR

A number of theoretical results have been shown concerning optimal strategies in market scoring rule markets. The first result, shown by [13], is a myopic honesty result: A risk-neutral trader who trades only once (or does not consider any future trades while making a report) will maximize her expected utility by reporting her true belief about the item.

The strategic situation is more complex when traders can trade repeatedly, and are non-myopic. Two specific kinds of non-myopic strategies that have been analyzed [3] are dubbed as *bluffing* and *delaying*. In a bluffing strategy, a trader first makes a trade that, with some probability, suggests information opposite to her true belief, so as to mislead other traders into reporting erroneous probabilities. In her next trade, she can then gain a profit by correcting the market price according to her true belief. Her total payoff will be the sum of the payoffs she earned from both trades. If she earns more from her second trade than she loses from her first trade, she gains a net profit. The extended game view of trade in a market also permits the delaying strategy: A trader with private information may choose to wait for other traders to report before revealing her private information. In comparison to the myopically optimal strategy, both bluffing and delaying have a negative effect on the speed of market convergence: The market price may not reflect the available information because one or more traders has either chosen to delay until later, or entered a report that is misleading.

In MSR-based markets, the profitability of these non-myopic strategies depends on the structure of traders' private information, i.e., the joint distribution of the signals they receive and the true outcome of the event. In particular, two natural distribution families have been studied: substitute and complementary signals.

In an information environment with substitute signals, private signals are independently distributed, *conditional on the true outcome*. For example, two people, A and B, try to predict if it is going to rain tomorrow. A tries to see if swallows fly low, and B uses the heuristic that "ring around the moon, rain is coming soon." Both A and B would receive private signals from "independent" sources about the weather tomorrow, though their signals are independent *conditional* on the current humidity of the air. In this case, the two signals are substitutes.

Another class of distributions involves signals that are (unconditionally) independent of each other. For example, firm A announces that at the end of the year each employee will receive a bonus if firm A's sales on both the East and West Coasts have met their targets. Employee E knows how the firm performed on the East Coast and employee W knows how it did on the West Coast. Assuming that the sales on the East and West Coasts are completely independent — knowing E's signals does not help one in guessing what W's signal is and vice versa. In this case, it can be shown that the signals are complementary: the predictive power of both traders' signals combined is greater than the sum of their individual predictive value.

When traders' private signals are substitutes, Chen *et al.* [3, 5] show that misleading non-myopic strategies are not profitable. Honest reporting of beliefs at the earliest opportunity is the only perfect Bayesian equilibrium. On the other hand, when traders' private signals are complements, Dimitrov and Sami [8, 3] show that non-myopic players can indeed profit from deviating from honest reporting by either bluffing or delaying.

To the best of our knowledge, there has been no experimental study on non-myopic strategic manipulation in MSR-based prediction markets. Experimental tests of these predictions can not only provide guidance to the designers of prediction markets, but also inform theory development in terms of suggesting future directions. Our study is the first one to test the theoretical predictions of this literature.

[1]See [24] and [6, p.139] for a discussion of various proper scoring rules.

2.3 Prior Experimental Work

There have been a number of studies conducted to measure the aggregative efficiency of prediction markets (see, for example, [21] and [22]). Apart from a few papers mentioned below, these have studied continuous double-auction markets or parimutuel markets, and not market scoring rule markets. We refer readers to the excellent literature review by Tziralis and Tatsiopoulos [25] for further information.

There have been experimental studies comparing the accuracy of the forecasts produced by prediction markets and other information aggregation methods. Ledyard *et al.* [17] study alternative forecasting techniques for combinatorial forecasting problems, and find that market scoring rules outperform all the alternatives studied. Graefe and Armstrong [12] found that structured information aggregation methods, including prediction markets, perform better than the unstructured information aggregation method, i.e. face-to-face meetings. Healy *et al.* [16] found that the relative performance of prediction markets to other alternative information aggregation methods, such as iterative polls, depends on the complexity of the information environment.

Our work is also related to prior work on manipulation in prediction markets. Hanson *et al.* [14] gave half of the subjects (the manipulators) incentives to manipulate the market price by pushing the price up. Other traders were informed of the existence of these manipulators and the direction in which they wanted to push price. [14] found the market price was robust to manipulations, because knowing what the manipulators were trying to do, other traders effectively counteracted their influences. Oprea *et al.* [19] extended [14]'s study by testing the influence of manipulators under the condition that all other traders only knew the existence of the manipulators but not the direction in which they push the price. They found that the traders still were able to counter-balance the manipulators' influence.

Our experiment differs from [14] and [19] in two aspects. First, we study internal manipulation — manipulations aimed at profiting within the same market — while [14] and [19] study external manipulation — manipulations aimed at profiting outside the market. To model the external manipulation, [14] and [19] gave the manipulators extra payments based on how successfully they influenced the market prices, in addition to their earnings as regular traders in the market. In our experiment, all the traders' payments are made as regular traders in the markets, even if they attempt to manipulate the market price. Second, our experimental prediction markets are based on a logarithmic market scoring rule, while [14] and [19] are based on double-auction markets. In our experimental design, the information distribution in the substitute treatment is consistent with the base model in [19], restricted to two traders.

3. EXPERIMENTAL DESIGN

Our experiment follows a between-subject design — each subject only participates in one treatment. We recruited 256 subjects who were all students at the University of Michigan. Before the experiment began, an experimenter read the instructions to all the subjects.[2] These instructions included a tutorial on the experimental market's software interface, each individual's payoff functions, and the information they would receive based on the treatment. The experimenter then administered a paper-based quiz to all the subjects and checked each subject's answers in person. No practice rounds were given before the data collection started. Communications among the subjects were strictly forbidden. After the experiment ended, each subject filled out a short post-experimental survey about the strategies they used.[3]

In all the treatments, subjects participate in the market via computer software. The experiment was programmed and conducted with the software z-Tree [10]. Each trader started with 200 units of experiment currency in each round. They were also informed that 133 units of experimental currency could be later exchanged for U.S.$1.[4] For each treatment, we ran 4 independent sessions to achieve sufficient repetitions. There were 8 subjects in each session, which consisted of 25 rounds. At the beginning of each round, the 8 subjects were randomly paired into 4 groups of 2 traders. There are two traders in each market. This is not a typical setup of a market: markets usually have more than two traders. Nonetheless, we chose to study two-trader markets due to their simplicity — it is a good starting point to observe basic market dynamics. Future work is needed to test our results in markets with a larger number of traders.

We use a 2 x 2 x 2 factorial design as shown in Table 1. The factors are: direct vs. indirect MSR, substitute vs. complementary private signals, and structured vs. unstructured trading order. The following sections contain details about these treatments.

3.1 Structured vs. Unstructured

Almost all the prediction markets used in practice have unstructured participation, in that people freely choose the timing and frequency of their trades. However, the theoretical analyses of strategic behaviors in prediction markets are largely based on the assumption of exogenous ordering: people are given opportunities to trade in an order predetermined by factors beyond their control [5, 8, 3]. In these models, they are not forced to trade at every opportunity they receive; however, any timing-game elements of the strategic interaction between traders is abstracted away. In particular, traders are modeled as knowing when they will receive future opportunities to trade, as well as knowing that other traders have had opportunities to trade between their trades. It is unclear to what extent people would behave differently in these two types of environments (exogenous vs. endogenous trading order). An empirical test that compares these two types of markets, while keeping all other factors constant, will test the validity of this assumption, and can help guide future theoretical development.

Apart from testing the theory, this comparison may also influence market design. In small scale prediction markets in practice, it is often possible to impose some structure on the trading order. In fact, the use of structured information aggregation tool is not an entirely new idea. For example, in the 1950s, the Delphi method was developed as a multiple-round survey to elicit expert forecasts [27]. In each round,

[2]The instructions can be downloaded at `http://www-personal.umich.edu/~rsami/papers/MarketExperiment/`.

[3]The survey questions are available at `http://www-personal.umich.edu/~rsami/papers/MarketExperiment/PostExperiment_Survey_Questions.pdf`.

[4]The average payment made to our subjects was \$42, with the minimum being \$28 and the maximum being \$52.

	Direct MSR		Indirect MSR	
	Substitutes	**Complements**	**Substitutes**	**Complements**
Structured	Str-Dir-Sub	Str-Dir-Comp	Str-Indi-Sub	Str-Indi-Comp
Unstructured	Unstr-Dir-Sub	Unstr-Dir-Comp	Unstr-Indi-Sub	Unstr-Indi-Comp

Table 1: Experimental Design

all participants are asked to provide their own forecasts and possibly comments. After each round, the aggregated forecasts are shown to all the participants, before they are asked to provide their revised forecasts again based on the aggregated forecasts of the group. The final forecasts are based on the aggregated forecasts in the final round. Our study is aimed at shedding light on whether structured prediction markets have advantages as well.

In treatments with a structured trading order, for each round we randomly determine the trading order between the two traders. The two traders then take turns to report their predictions; they may, of course, leave the previous report unchanged if they wish. In total, each trader has three turns to report. When it is a trader's turn, she has 30 seconds to make a decision. In treatments with an unstructured trading order, all traders can choose when to trade during the two-minute window in which the market is open.

3.2 Direct vs. Indirect MSR

We implement our MSR market using the logarithmic scoring rule, i.e., $s_i(r) = log(r_i)$, and call such type of market a *logarithmic market scoring rule* (LMSR) market. All the subjects report their prediction of the probability of the *black* ball being drawn in percentages.[5] Individual m's payoff from her report r_m is $200 \times (log_{10}(r_m) - log_{10}(r_{m-1}))$, where r_{m-1} is individual $m-1$'s report. Note that we used a scaler 200 to adjust the extent to which a subject can influence the market price. The initial market prediction is set to 50 (%). All the transactions in a market are displayed in real time to both participants.

In an indirect MSR market, subjects trade securities, each of which is based on a possible outcome of the random event. There are two securities in the markets, black and white, each paying one unit of our experiment currency if the corresponding outcome is realized, and zero units of experiment currency otherwise. The underlying market scoring rule and its parameters are exactly the same as those used in the direct MSR markets. To simplify the interface of an indirect MSR market, we only support trades in multiples of 10 and 50 shares. The exact prices of the shares and the new price after the transaction are shown to the subjects in real time. Restricting the number of shares per transaction in the indirect MSR markets might have an impact on the accuracy of the market predictions, but we argue that such an impact is likely small. Figure 1 illustrates how much increase in the market price can a purchase of 10 shares cause in an indirect MSR market. The highest price increase per 10-share purchase, 2.8 cents, occurs when the current price is 50 cents. As the current market price move away from 50 cents, the impact of a 10-share purchase decreases.

[5]Theoretically the range of probability should be 0% to 100%. But as the logarithmic function is undefined at 0, we restricted the probability predictions as integers in the range of $[1, 99]$.

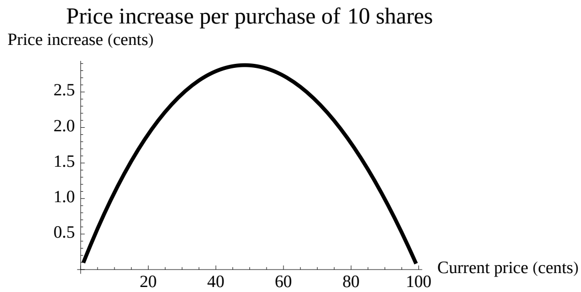

Figure 1: The price increase per purchase of 10 shares in an indirect MSR market.

3.3 Substitutes vs. Complements

The information signals that traders see are generated according to a pre-specified distribution. We designed two information distribution treatments corresponding to the substitute and complement conditions as described in Section 2.2. In each treatment, all subjects try to predict the outcome of a hypothetical random event, based on their private information given by the experimenter, while sharing a common prior belief about the distribution of the outcomes.

For the substitutes treatment, we use the same substitutes signals as used in [19]. At the beginning of each round, the computer randomly draws a black or white ball with equal probability. Once the round begins, each subject receives a private signal, either a "+" or a "-". The signal each subject receives depends on the color of the ball drawn at the beginning of the round. If the black ball was drawn, the signal will be a "+" with a 2/3 chance; and if the white ball was drawn, the signal will be a "-" with a 2/3 chance. At the end of each round, the color of the ball drawn is revealed, based on which all the subjects receive their individual payments.

In order to compare traders' behaviors in substitutes and complements markets, we specify an information condition with complementary signals such that the two are comparable. We impose the following two criteria on the complement signal structure to achieve a fair comparison:

C.1 The prior distributions of the random outcomes are the same under both environments. That is, the color of the ball drawn is black or white with equal probability.

C.2 The expected earnings of all the subjects are the same under both information conditions.

Criterion C. 1 ensures that the subjects in both treatments have the same prior belief about the security, to rule out confounds that people may behave differently when dealing with different prior probabilities. Criterion C. 2 implies that the value of the signals are equal. Specifically, if one person observes both subjects' private signals, in expectation, he would earn the same payoffs in both environments.[6]

[6]Since we are using the logarithmic scoring rule, the fact

# of "+"s	# of "-"'s	Prob Black	Prob White
0	2	19%	81%
1	1	50%	50%
2	0	81%	19%

Table 2: Mapping signals to the probability of the black or white ball being drawn

We found the following complement signal structure that satisfies both criteria C. 1 and C. 2. Once the round begins, each subject receives a signal, a "+" or a "-" randomly drawn with equal probability. At the end of each round, a color (black or white) is randomly drawn by the computer, depending on the numbers of each signal that everyone together has received. How the signals determine the probability that a black ball will be drawn is shown in Table 2. For example, if there are 0 "+"'s and 2 "-"'s (there are only two traders in the market), there is a 19% chance that the black ball will be drawn, and an 81% chance that a white ball will be drawn. Note that, due to the different generating processes, the posterior probability of drawing a black ball given, say, two "+" signals, is slightly different in the complements treatment. It follows that, conditioned on getting two "+" signals, there is a different total expected profit from ideal aggregation in the substitutes and complements treatments. However, the probability of getting two "+" signals is also different, and by design, these two factors balance so that the expected total profit is the same in the two treatments.

3.4 Analysis Metrics and Hypotheses

We define *posterior efficient price* (PEP) as the prediction a perfect Bayesian who has observed all the signals in a round and has the correct prior belief would have reached. In theory, a LMSR market with perfect information aggregation would converge to the PEP [13]. We will use the PEP as a benchmark to measure the accuracy of the forecasts made by our experimental prediction markets. We measure the market prediction accuracy as the mean squared error (MSE) of the market closing price to the PEP.

An alternative, and perhaps more natural, metric would have measured the distance of the final price from the actual realized outcome (i.e., whether the ball is black), or the correlation between the price and the realized outcome. In the long run, this metric would convey exactly the same information as the distance from the PEP. However, the actual outcome is subject to an additional layer of randomness, reflecting the distribution of the final outcome conditioned on the signal-pair of the traders. As we seek to measure the performance of the market in aggregating available information rather than serendipitously matching the true outcome, the PEP is a better comparison point: it yields the same long-run average performance, but for a finite number of rounds, the measure of the distance from the true outcome is noisier. In using the PEP, we are merely taking advantage of having controlled experiments rather than field trials.

With the analysis metric defined, we summarize our hypotheses and research questions below. First, we do not have any theory in predicting the effects of trading ordering on the performance of the prediction markets. Thus, we pose it as an open question:

- Do structured markets produce more accurate predictions than unstructured markets?

Second, since the direct and indirect MSRs are strategically equivalent (see Section 2.1), theoretically, varying the implementation of the MSR should not affect the accuracy of the market prediction.

HYPOTHESIS 1. *The MSEs of the forecasts produced by direct and indirect MSR markets are the same.*

Third, theory predicts that there will be more delayed trading and bluffing in complements markets. Hence we have the following two hypotheses on traders' behavior:

HYPOTHESIS 2. *There is more bluffing in complements markets than in substitutes markets.*

HYPOTHESIS 3. *There are more delayed trades in complements markets than in substitutes markets.*

And at an aggregated level, behaviors predicted in Hypotheses 2 and 3 would lead to the following hypothesis:

HYPOTHESIS 4. *Market prediction converges faster in substitutes markets than in complements markets.*

4. RESULTS

4.1 Market Performance

Table 3 contains the MSEs of the market closing prices of all the treatments. Note that the lower the MSEs, the better the market prediction accuracy.

Structured vs. Unstructured. Treating the four sessions of each treatment as independent data points, we conducted permutation tests to compare structured with unstructured markets while holding other conditions constant. Table 4 contains the results of the relevant tests. All the tests have the null hypotheses that the MSEs of the prices in both treatments are equal. For example, row 1 reads "under the condition of direct MSR and substitute private signals, the mean squared errors of the market closing prices in unstructured markets are higher than those in structured markets. The result is statistically significant at the 5% level (p-value = 0.028)."

We found that structured markets perform better than unstructured markets: under three out of the four conditions (rows highlighted in bold in Table 4) the difference is statistically significant at the 5% level. Under the fourth condition, indirect MSR and complementary signals, although the difference is not statistically significant, it is in the expected direction: see column 4 in Table 3.

The differences between the MSEs of structured and unstructured markets are fairly large, as shown in Table 3. For example, in markets with direct MSR and substitutes signals, the MSE of the structured markets, 265, is significantly lower than that of the unstructured markets, 413 (see column 1 in Table 3). These results suggest that the

that the two signals have the same value has an interpretation in information theory: in both settings, the two signals lead to the same reduction in the entropy (uncertainty) of the forecasted event.

	Direct MSR		Indirect MSR	
	Substitutes	**Complements**	**Substitutes**	**Complements**
Structured	265	289	267	345
Unstructured	413	385	446	386

Table 3: Market prediction accuracy comparisons

Condition	Alternative Hypotheses	P-value
Dir, Sub	**Unstr > Str**	**0.028**
Dir, Comp	**Unstr > Str**	**0.014**
Indi, Sub	**Unstr > Str**	**0.01**
Indi, Comp	Unstr > Str	0.27
Number of independent obs. per treatment		4

Table 4: Results of permutation tests comparing the MSEs of the market prices between structured and unstructured markets.

Condition	Alternative Hypotheses	P-value
Unstr, Sub	Direct $\neq$ Indirect	0.2
Unstr, Comp	Direct $\neq$ Indirect	0.8
Str, Sub	Direct $\neq$ Indirect	0.47
Str, Comp	Direct $\neq$ Indirect	0.21
Number of independent obs. per treatment: 4		

Table 5: Results of permutation tests comparing the MSEs of the market prices between markets with direct and indirect mechanisms.

additional structure provided by the fixed trading order improves traders' ability to interpret others' signals from their trades, and combine it with their own information. These results are consistent with Graefe and Armstrong's finding that structured information aggregation methods work better than unstructured ones [12].

Direct vs. Indirect MSR. The test results related to comparing the two trading mechanisms are listed in Table 5, following the same format as in Table 4. Based on the results, we cannot rule out Hypothesis 1. We found no difference in the market prediction accuracy between the two mechanisms under all four possible combinations of the two information conditions and the two trading orders (all the tests in Table 5 are not statistically significant). As shown in Table 3, the actual MSEs of the market closing prices in these two mechanisms are very similar. In particular, in structured substitutes markets, the difference is 1 (265 vs. 267); in unstructured complements markets it is also 1 (385 vs. 386).

This result addresses an active debate in the prediction market research community about which mechanism is more effective. Within our controlled environment, there was no significant difference between the overall aggregative performance of the prediction market given the differing mechanisms. Further research needs to be carried out to determine if this is validated in field settings as well, as traders may not read and follow instructions as carefully in a field setting.

Substitutes vs. Complements. In markets with a structured trading order, we found evidence supporting Hypothesis 4 — market prediction converges faster in substitutes markets than in complements markets. However this hypothesis is not supported in markets with an unstructured trading order.

Table 6 contains the results of the relevant tests. To establish a convergence result, we measured the MSE of the market price at two point in time: first right after both traders each had one chance to trade (marked as "early" in Table 6) and second when the market closes. We then compared the MSEs of the prices of a market at these two points across treatments. In structured direct MSR markets, we found no statistically significant difference between the MSEs of the market closing prices between the substitutes and complements markets (row 1 in Table 6, p-value = 0.23, two-sided test; and see row 3, p-value=0.16, two-sided test).

However when comparing the MSEs of the market prices right after both traders had had one turn to trade, we found the MSEs of the complements markets higher than those of the substitutes market (see row 2 in Table 6, p-value = 0.057, one-sided test; and row 4, p-value = 0.08, one-sided test). We did not find a similar price convergence pattern in unstructured markets (see rows 5 - 8 in Table 6). One possible reason for this is that, in the unstructured markets, the increased difficulty of inferring others' signals (as evidenced by the worse overall aggregation performance) makes it more difficult to successfully execute a bluff-and-correct attack, or to exploit the greater value of one's information signal given the other trader's revealed information.

4.2 Strategic Behaviors

4.2.1 Bluffing

In structured direct MSR markets, we found support for Hypothesis 2 — there is more bluffing in complements markets than in substitutes markets. There were more groups whose first trades were inconsistent with their traders' private signals in the complements markets (23 out of 100 round/groups on average) than in the substitutes markets (15 out of 100 round/groups on average). Again we used a permutation test to compare the mean number of round/groups with dishonest first trades in both treatments, while treating each of the four sessions in each treatment as an independent data point. The test result shows that there was a statis-

Condition	**Alternative Hypotheses**	**P-value**
Str, Dir	Comp $\neq$ Sub	0.23
	Comp (early) $>$ Sub (early)	0.057
Str, Indi	Comp $\neq$ Sub	0.16
	Comp (early) $>$ Sub (early)	0.08
Unstr, Dir	Comp $\neq$ Sub	0.63
	Comp (early) $\neq$ Sub (early)	0.19
Unstr, Indi	Comp $\neq$ Sub	0.67
	Comp (early) $\neq$ Sub (early)	0.77
Number of independent obs. per treatment: 4		

Table 6: Results of permutation tests comparing the MSEs of the market prices between substitutes and complements markets.

tically significant higher number of dishonest first trades in complements markets than in substitutes markets (p-value = 0.056, one-sided test). In the structured indirect market, again we saw a greater number of first trades contrary to the trader's information in the complements treatment than in the substitutes treatment (28 per 100 vs. 21 per 100), but this result was of marginal significance (p-value = 0.13, one-sided test). In the unstructured markets (unstructured direct and indirect MSR), however, the number of round/groups with dishonest first trades was not significantly different between substitutes and complements markets.

Our survey data corroborates this finding. Table 7 shows the results of three linear regressions predicting the subjects' level of agreement to the statement: "To maximize my own profit, the best strategy is to report honestly according to my private information." The dependent variable takes five possible values in the set $\{-2, -1, 0, 1, 2\}$, with -2 indicating strongly disagree, and 2 indicating strongly agree. The independent variables include the dummy variables for the three treatment respectively. For example, *structured* takes a value of 1 if the subject was in a treatment with structured trading order. In regression (2) we also included three interaction terms among the three treatments. Regression (1) and (2) were run on a dataset containing data from all the sessions, and regression (3) was run on a subset of the data which includes only the treatments with structured trading order and direct MSR.

Our survey data suggest that in markets with complementary signals, traders are less likely to think honest trading is the optimal strategy. Regression (1) shows that across all the treatments, trading with others who have complementary signals leads to a 0.23 point (out of 4) reduction in the level of agreement to honest trading as the best strategy. In regression (2) we found that such a reduction is more pronounced in structured markets — the size of the reduction is 0.40, bigger than the main effect estimated in regression (1) (0.23). When we focus on the structured direct MSR markets in regression (3), we found an even bigger effect, a 0.56 point reduction in the subjects' level of agreement in

Dependent variable: level of agreement to honest trading as the best strategy			
	(1)	(2)	(3)
Structured	-0.062	0.125	
	(0.126)	(0.180)	
Complements	-0.234	0.094	-0.562
	(0.126)*	(0.222)	(0.237)**
Direct	-0.016	0.094	
	(0.126)	(0.226)	
Str $\times$ Comp		-0.406	
		(0.237)*	
Str $\times$ Direct		0.031	
		(0.237)	
Comp $\times$ Direct		-0.250	
		(0.237)	
Constant	0.250	0.094	0.344
	(0.133)*	(0.156)	(0.129)**
Observations	256	256	64
R-squared	0.014	0.028	0.077

Table 7: Linear regressions predicting the level agreement to honest trading as the best strategy
Notes: (i) Each column corresponds to a different regression, as detailed in the text. (ii) Standard errors are clustered at the session level. (iii) Standard errors in parentheses. * significant at 10%; ** significant at 5%.

honest trading being the best strategy.

Now that we have observed some bluffing in both substitutes and complements markets, is bluffing profitable given the traders' reactions to each others' trades? In theory, we expect the incremental profit due to bluffing to be negative in the substitutes treatment. In the complements treatment, we expect the incremental profitability of bluffing to be 0 in equilibrium: if it were positive or negative, the bluffer could profitably increase or decrease the probability of bluffing. To answer this question, we conducted multivariate linear regressions on a panel dataset based on each subject's behavior in each round. To identify potential bluffing, for each group of each round, we only took the subject who trades first. Since for each treatment there are 4 groups in each round, 25 rounds in each session, and 4 sessions in total, we have a sample size of 400.[7]

Our dependent variable is the expected profit of the subject for the round given the trades she made, weighed by the posterior distribution of the outcomes. For example, suppose in a round with substitute signals, traders A and B both receive a "+" signal. Subject A makes a trade that changes the current market prediction from 50% to 65%. Her payoff will be $200 \times (log_{10}(0.65) - log_{10}(0.5)) = 23$ if the outcome is black, and $200 \times (log_{10}(0.35) - log_{10}(0.5)) = -31$ if the outcome is white. Given two "+" signals in the market, the posterior probability of the outcome being black is

$$\frac{\frac{2}{3} \times \frac{2}{3} \times \frac{1}{2}}{\frac{2}{3} \times \frac{2}{3} \times \frac{1}{2} + \frac{1}{3} \times \frac{1}{3} \times \frac{1}{2}} = 0.8 \tag{3}$$

and the posterior probability of the outcome being white is 0.2. Subject A's expected profit from this trade then is $0.8 \times 23 + 0.2 \times (-31) = 12.24$. If subject A makes multiple

[7] In the complements markets in some rounds neither of the traders traded, so there are only 388 data points.

trades in this round, her expected profit in this round would be the sum of her expected profit of all the trades she made in this round.

Our independent variables are the following:

- Gain — Dummy variable. It equals 1 if there is a profit to be gained in the market. Gain = 1 if $PEP \neq 50$ and Gain = 0 if $PEP = 50$.
- Bluff — Dummy variable. It equals 1 if the trade is inconsistent with the subject's private signal.
- BigBluff — Dummy variable. It equals 1 if the trade is a bluff and if it moves the current market price by more than 10 points.
- SmallBluff — Dummy variable. It equals 1 if the trade is a bluff and if it moves the current market price by less than 10 points.

We used a fixed effect model to account for repeated observations of each subject. The standard errors are clustered at the session level to adjust for intra-session influences among the subjects. Using fixed effects allows us to control for individual effects, which might be correlated with some of our independent variables, e.g., Bluff, BigBluff, and SmallBluff. The results are summarized in Table 8. For each market, we constructed models (1) and (2) to explore the main effects of bluffing and the effects of bluffing by different amounts, i.e., a small bluff (represented by variable SmallBluff) and a big bluff (represented by variable BigBluff).

In the substitutes markets, as an aggregate effect, we did not find a statistically significant effect of bluffing on expected profit of the round — we cannot rule out the hypothesis that the coefficient on Bluff, -2.349, is different from 0 by chance. However this does not mean that bluffing is not profitable in substitutes markets. When we break the Bluff variable down into BigBluff and SmallBluff, we found significant effects of both. If one bluffs by a large amount, her profit decreases by 10.8 points on average. But if she chooses to bluff by a small amount, she can earn about 10.92 points. Both coefficients are statistically significant at the 5% level. This result contradicts what theory predicts — if all players are perfectly rational and risk neutral, it is not profitable to bluff in substitute signal environments. However, in practice this is not true. Our subjects discovered that bluffing by small amounts is profitable. A possible explanation is that when bluffing, perhaps one does not have to move the price by very much to convince her opponent that she has a signal that is opposite to her true signal.

Contrary to expectations, our result from model Comp(1) shows that, on average a bluff is associated with a 7.8 point *loss* (statistically significant at 10% level). There are multiple potential explanations to this outcome. It could be that bluffers are bluffing too often, such that other traders stopped believing the information revealed. Hence bluffers are less likely to successfully execute their strategy. Or, it could be that traders are over-confident of their own private signals, such that they do not fully incorporate other traders' signals when they trade. In this case, bluffers will make a smaller profit than would be expected in theory. It could also be that, as a complex strategy, bluffing is more sensitive to errors in execution. Repeating the same analysis with big and small amount of bluffing, we did not find significant results in model Comp(2).

Dependent variable: the actual first move of the second mover	
Complements	0.217
	(0.122)*
Constant	2.327
	(0.117)***
Observations	734
Number of Subjects	64

Table 9: Linear regression predicting the actual first move of the second mover

Notes: 1) We use a random effect model and standard errors are clustered at the session level. 2) Standard errors in parentheses. * significant at 10%; *** significant at 1%

4.2.2 Delayed Trading

Again in the structured direct MSR markets, we found evidence supporting Hypothesis 3 — there are more delayed trades in complements markets than in substitutes markets. We restrict our analysis to the delaying behavior of the second mover — the trader who is randomly selected by the computer to move second in each round. This is because in our experimental setting, it is unclear what the first mover would gain by delaying her trades. There are six possible moves in each round. The first mover has the opportunity to trade during moves 1, 3, and 5. The second mover has the opportunity to trade during moves 2, 4, and 6. Since the last trading opportunity is given to the second mover, if the first mover is to reveal her signal, she has to do so by the end of move 5.

For the second mover, we define a variable called ActualFirstMove, which is the move during which the second mover makes her first trade. The mean ActualFirstMove in the complements markets is 2.51, higher than the mean in the substitutes markets, i.e., 2.3. Treating each session as an independent data point, we conducted a permutation test to compare the mean ActualFirstMove between sessions of the complements treatment and sessions of the substitutes treatment. The result is marginally significant: p-value = 0.099, one-sided test. To test the robustness of this finding, we conducted a linear random effect regression on a dataset consisting of all second mover's first trades in each group of each round. The dependent variable is still the ActualFirstMove, and the independent variable, Complements, is a dummy variable which takes a value of 1 if the treatment is complements and 0 otherwise. A random effect model is appropriate here because the independent variable is exogenous — it is a treatment applied to randomized subjects. The result is shown in Table 9.

The coefficient on the complements variable is 0.217 (p-value = 0.07), and it is statistically significant at the 10% level. This result is consistent with what theory predicts — there are slightly more delayed trading in complements markets than in substitutes markets.

5. CONCLUSION

We conducted human-subject experiments to analyze the performance of variants of market scoring rule-based prediction markets, under differing information conditions.

We found that markets with structured trading orders provide better predictions than those with unstructured trad-

Dependent variable: the subject's expected profit of the round				
	Sub(1)	Sub(2)	Comp(1)	Comp(2)
Gain	17.082	17.902	17.457	17.45
	(1.959)***	(1.822)***	(3.585)**	(3.606)**
Bluff	-2.349		-7.793	
	(-3.459)		(2.828)*	
BigBluff		-10.798		-8.101
		(2.237)**		(-3.644)
SmallBluff		10.916		-7.206
		(3.362)**		-4.27
Constant	192.441	192.011	192.615	192.61
	(0.479)***	(0.492)***	(2.008)***	(2.006)***
Observations	400	400	388	388
Number of Subjects	32	32	32	32
R-squared	0.126	0.152	0.149	0.149

Table 8: Linear regression predicting expected profit

Notes: 1) Fixed effect model and standard errors are clustered at the session level. 2) Standard errors in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%

ing orders. This suggests a possible modification of prediction markets that may be beneficial when feasible, with small groups of traders. We also compared the performance of the direct and indirect MSR markets. We found no significant difference between the two forms.

Our experiments also enabled us to test theoretical predictions of strategic behavior in prediction markets. We compared the performance of markets under two different signal distribution conditions: when traders' signals are substitutes and when they are complements. In markets with a structured trading order, we found evidence supporting the theoretical prediction that there will be more manipulative behaviors, i.e., delaying and bluffing, in the complement signal markets than in the substitute signal markets. This was confirmed in our subject surveys, indicating that the subject group could perceive the strategic difference between the two environments. Interestingly, the theoretical predictions were not borne out in treatments without a structured order of trading opportunities. This result suggests that future theoretical work on the strategic behavior of prediction market traders will need to take into account the endogenous trading order typical of deployed prediction markets.

Our results suggest two important directions for future research. Firstly, our experiments were conducted in two-trader markets so that we could closely test the theoretical predictions and make cleaner inferences about trader behavior; it is important to confirm the validity of our conclusions with larger group experiments. Secondly, field experiments will complement our lab experiments comparing direct and indirect MSR: the effect of variations in interface may be more pronounced when users are not provided training in a controlled laboratory environment.

6. REFERENCES

[1] H. Berg and T. A. Proebsting. Hanson's automated market maker. *The Journal of Prediction Markets*, pages 45–49, Jan 2009.

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

[3] Y. Chen, S. Dimitrov, R. Sami, D. Reeves, D. Pennock, R. Hanson, L. Fortnow, and R. Gonen. Gaming prediction markets: Equilibrium strategies with a market maker. *Algorithmica*, 2009.

[4] Y. Chen and D. M. Pennock. A utility framework for bounded-loss market makers. In *Proceedings of the 23rd Conference on Uncertainty in Artificial Intelligence (UAI 2007)*, pages 49–56, Vancouver, BC Canada, 2007.

[5] Y. Chen, D. M. Reeves, D. M. Pennock, R. D. Hanson, L. Fortnow, and R. Gonen. Bluffing and strategic reticence in prediction markets. In *Lecture Notes in Computer Science, The Third Workshop on Internet and Network Economics (WINE 2007)*, pages 70–81, 2007.

[6] R. M. Cooke. *Experts in Uncertainty: Opinion and Subjective Probability in Science*. Oxford University Press, New York, 1991.

[7] B. Cowgill, J. Wolfers, and E. Zitzewitz. Using prediction markets to track information flow: Evidence from google, 2008.

[8] S. Dimitrov and R. Sami. Non-myopic strategies in prediction markets. In *Proceedings of the Nineth ACM Conference on Electronic Commerce (EC'08)*, New York, July 2008. ACM.

[9] E. F. Fama. Efficient capital market: A review of theory and empirical work. *Journal of Finance*, 25:383–417, 1970.

[10] U. Fischbacher. z-tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2):171–178, 2007.

[11] R. Forsythe, F. Nelson, G. R. Neumann, and J. Wright. Anatomy of an experimental political stock market. *American Economic Review*, 82(5), 1992.

[12] A. Graefe and J. S. Armstrong. Comparing face-to-face meetings, nominal groups, delphi and prediction markets on an estimation task. Working paper.

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

[14] R. Hanson, R. Oprea, and D. Porter. Information aggregation and manipulation in an experimental market. *Journal of Economic Behavior and*

Organization, Jan 2006.
[15] F. A. Hayek. The use of knowledge in society. *The American Economic Review*, 35:519 – 530, 1945.
[16] P. J. Healy, J. O. Ledyard, S. Linardi, and J. R. Lowery. Information aggregation mechanisms and environmental complexity. Working paper, 2009.
[17] J. Ledyard, R. Hanson, and T. Ishikida. An experimental test of combinatorial information markets. *Journal of Economic Behavior and Organization*, 69(2):182–189, 2009.
[18] J. F. Muth. Rational expectations and the theory of price movements. *Econometrica*, 29:315–335, 1961.
[19] R. Oprea, D. Porter, C. Hibbert, R. Hanson, and D. Tila. Can manipulators mislead market observers? *Working paper*, 2007.
[20] D. Pennock. Implementing Hanson's Market Maker, 2006. Blog post, available at `http://blog.oddhead.com/2006/10/30/implementing-hansons-market-maker/`.
[21] C. R. Plott and S. Sunder. Efficiency of experimental security markets with insider information: an application of rational-expectations models. *The Journal of Political Economy*, 90:663–698, 1982.
[22] C. R. Plott and S. Sunder. Rational expectations and the aggregation of diverse information in laboratory security markets. *Econometric*, 56(5):1085–1118, 1988.
[23] R. Radner. Rational expectations equilibrium: Generic existence and the information revealed in prices. *Econometrica*, 47(3):655–678, 1979.
[24] R. Selten. Axiomatic chracterization of the quadratic scoring rule. *Experimental Economics*, 1(1):43–62, June 1998.
[25] G. Tziralis and I. Tatsiopoulos. Prediction markets: An extended literature review. *Journal of Prediction Markets*, 1(1):75–91, 2007.
[26] J. Wolfers and E. Zitzewitz. Prediction markets. *Journal of Economic Perspectives*, 18(2):107–126, 2004.
[27] F. Woudenberg. An evaluation of delphi. *Technological Forecasting and Social Change*, 40:131–150, 1991.

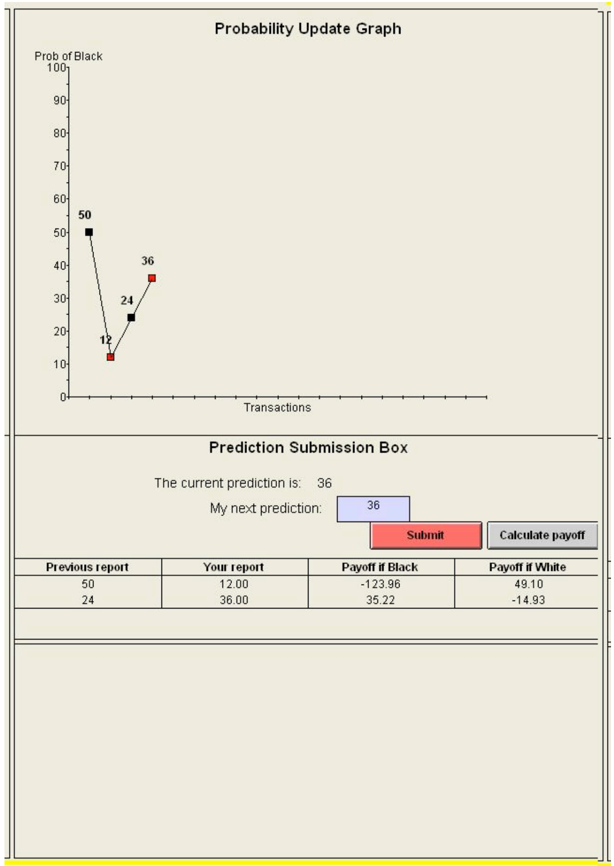

Figure 2: The trading interface of the direct MSR markets

APPENDIX

A. COMPUTER INTERFACES

Figures 2 and 3 show the trading interfaces used in our direct and indirect MSR markets respectively. They each contain two boxes: a probability or price update graph that displays the market activities in real time and a prediction or transaction submission box in which they enter their trades. With the direct MSR as shown in Figure 2, all a trader has to do is to report her prediction in the box next to "My next prediction". In indirect MSR markets as shown in Figure 3, traders look at the current prices for the black and white securities and make their trading decisions by clicking the trading buttons provided to them.

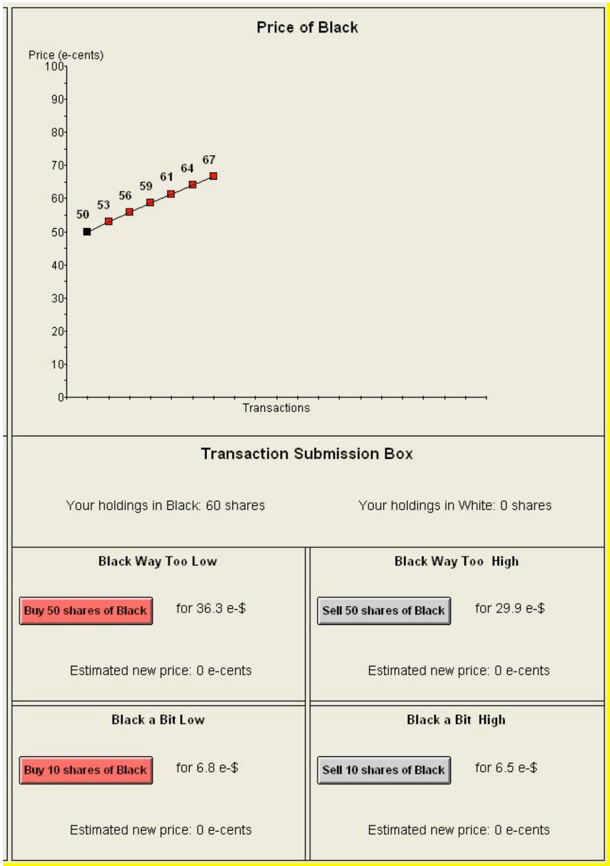

Figure 3: The trading interface of the indirect MSR markets