Prediction Market Alternatives for Complex Environments*

Paul J. Healy
The Ohio State University†

John O. Ledyard
California Institute of Technology‡

Sera Linardi
California Institute of Technology§

J. Richard Lowery
University of Texas at Austin¶

Abstract

Prediction markets have proven successful in large-scale applications such as elections and sporting events. Consequently, several large corporations have adopted prediction markets for smaller-scale internal applications where information may be complex and the number of traders small. Using laboratory experiments we test the performance of the standard prediction market in complex environments with few traders and compare it to three alternative mechanisms. When information is complex we find that an incentivized iterated poll (or Delphi method) out-performs the prediction market mechanism. We present four behavioral observations that may explain why the poll performs better in these settings.

JEL: C72, C91, D82

Keywords: Information aggregation, prediction markets, mechanism design.

*The authors thank Rachel Croson, John Fountain, Rick Green, Joel Grus, Glenn Harrison, Ernan Haruvy, Shimon Kogan, Tony Kwasnica, Sherry Li, Bryan Routledge, and Justin Wolfers for helpful comments and conversations and Andrej Svorencik for useful research assistance.

†Corresponding author. 1945 North High Street, Columbus, Ohio, 43210. Phone: (614) 247-8876. Fax: (614) 292-3906. Email: healy.52@osu.edu.

‡1200 East California Boulevard, Mail Code 228-77, Pasadena, California, 91125. Email: jledyard@hss.catlech.edu.

§1200 East California Boulevard, Mail Code 228-77, Pasadena, California, 91125. Email: slinardi@hss.caltech.edu.

¶Finance Department, McCombs School of Business, 1 University Station, Austin, TX 78712. Email: Richard.Lowery@mccombs.utexas.edu.

1 Introduction

In large scale applications prediction markets have proven successful at predicting future outcomes. The Iowa Electronic Market and the TradeSports-InTrade exchanges have out-performed national polls in predicting winners of political elections (Berg et al., 2003; Wolfers and Zitzewitz, 2004), as did an underground political betting market in the late nineteenth and early twentieth centuries (Rhode and Strumpf, 2004). Even markets with 'play' money incentives such as the Hollywood Stock Exchange and the NewsFutures World News Exchange perform as well as real-money exchanges in predictive accuracy (Servan-Schreiber et al., 2004; Rosenbloom and Notz, 2006).[1]

These successes in large-scale applications have led many large corporations—including Google, Hewlett-Packard and Intel—to adopt prediction markets for smaller-scale internal applications such as predicting future sales volumes of a particular product (Chen and Plott, 2002; Hopman, 2007; Cowgill et al., 2009).[2] It is not obvious, however, that the successes observed in large-scale settings will extend to smaller applications. In a corporate setting private information may be too weak, too widely dispersed, or too difficult to interpret for markets to aggregate information accurately. Management may want to collect information on variables that are correlated along several dimensions. Although prediction markets should be capable of aggregating this information in theory, it may be difficult in practice when traders face cognitive constraints and uncertainty about the rationality of others. These problems are exacerbated by the use of a relatively small numbers of traders since individuals may have market power that prevents convergence to the perfectly competitive outcome and therefore hinders the potential for information aggregation. In short, the assumptions of rational expectations and perfectly competitive markets seem at odds with the corporate environments where these markets are now being applied.

Given these potential difficulties there may be alternative information aggregation mechanisms that would outperform the standard double-auction prediction market in smaller-scale settings with complex or disperse information. For example, a variant of the Delphi method—where informed parties make predictions, learn each others' predictions, and then revise their own predictions—could be used to aggregate individuals' beliefs or private information, or a parimutuel-style betting market could be run to estimate the odds of certain future events.

In this paper we employ a behavioral mechanism design methodology, using laboratory experiments to test the performance of the prediction market mechanism in environments with a small numbers of traders and complex information structures. We extend our analysis by comparing market performance in an environment with a moderately complex information structure with only one true-false event to a second environment with a highly complex information structure featuring three correlated true-false events. We then compare the market's performance in these

[1] Rosenbloom and Notz (2006) do find that TradeSports significantly out-performs NewsFutures for some bundles of commodities and with enough data, but most tests cannot reject the null hypothesis of equal accuracy.

[2] Cowgill et al. (2009) identify at least twenty-one sizeable corporations that have used prediction markets.

environments to the performances of three alternative mechanisms for aggregating information. Specifically, we compare the standard double auction mechanism to an iterated polling mechanism, a parimutuel betting mechanism, and a synthetic 'market scoring rule' developed by Hanson (2003). By exploring the performance of these mechanisms in the laboratory we can gain an understanding about the domains on which each succeeds or fails and we can also acquire some insight into the reasons *why* some mechanisms out-perform others by understanding how agents' behavior is affected by the details of the mechanism. Ultimately, insights such as these serve as inputs into the 'behavioral' mechanism design process, providing guidance to practitioners hoping to design information aggregation mechanisms for use in these small-scale settings.

We find that the prediction market mechanism performs relatively well in an environment with a simple information structure involving one true-false event. In contrast, when the information structure becomes complex—with three correlated events and eight securities—the iterative poll performs the best and the standard prediction market the worst. Thus, we find strong support for the claim that the complexity of the environment interacts with the details of the mechanism. For example, traders in the prediction market with eight securities tend to focus attention on a small subset of the eight markets, causing severe mispricing in the remaining markets. The iterated poll avoids this issue by requiring players to announce beliefs about all eight states of the world simultaneously. In this way the design of the mechanism can be used to overcome natural behavioral biases that hinder information aggregation.

Our results suggest the following guidance for practitioners: In simple settings with a large number of traders relative to the number of items being predicted we suggest using the standard prediction market mechanism. When the number of items being predicted grows large (or the number of traders is small) and when those items are correlated we suggest the incentivized iterated poll instead. One downside of the iterated poll is that it requires subsidy payments from the institution running the mechanism; the size of these subsidies is limited, however, since we suggest using this mechanism only when the number of traders is relatively small. For larger environments the unsubsidized prediction market mechanism is preferable. The parimutuel mechanism is less desirable because it appears to suffer from no-trade outcomes where agents prefer to opt out of the mechanism entirely, as is predicted by the no-trade theorem of Milgrom and Stokey (1982). We do not suggest the market scoring rule (MSR) because it tends to suffer from informational 'mirages' where the mechanism leans toward completely *incorrect* predictions.

We follow our main results on mechanism performance with an analysis of four behavioral observations that we believe are related to the failure of the market mechanism and the success of the iterated poll in the complex setting. First, we see several apparent attempts at market manipulation in the prediction market mechanism and in the poll, but very few in the iterated poll and MSR. This is expected in the iterated poll; all players receive the same earnings and therefore have no clear incentive to manipulate their opponents' information. Second, total payments in the poll and MSR are subsidized by the mechanism designer, so all traders have an incentive to participate actively. Third, traders in the market appear to focus attention on only a subset of the securities—a heuristic that is impossible in the poll since it requires each trader to submit an entire probability distribution. Finally, an aberrant or confused trader can significantly affect final outcomes in the market, parimutuel, or MSR, but not in the poll because the poll takes the average of traders' reports as the predictive distribution.

These four observations allow us to extrapolate our results beyond the four mechanisms tested and to guide the design of future mechanisms. For example, a designer of other mechanisms for information aggregation should consider those with aligned incentives, subsidized total payments (if feasible), a focus on entire probability distributions, and minimal reliance on any one individual's report. Our results also inform economic theory: theories of market equilibration should take into account the tendency for traders to manipulate others or to focus attention (or coordinate) on a subset of available markets. As such theories are developed and refined they could then be used to inform the design of additional mechanisms.

This paper extends past work on market efficiency and information aggregation. The number of traders in the market is often cited as a factor that affects the degree of efficiency and information aggregation, though the effect likely depends on the proportion of traders who hold valuable information. Clearly, the presence of additional informed traders increases the amount of information that is available to aggregate, but the effect of additional noise traders with no private information is unclear. DeLong et al. (1990) argue that noise traders' uninformed trades can reduce the informational content of market prices and damage market efficiency, while Kyle (1985) shows how noise traders can provide profit opportunities for informed traders, inducing them to make larger trades and invest more resources—physical or cognitive—in the acquisition and integration of information. Empirical evidence on the issue is mixed; volume is positively correlated with accuracy in the Iowa Electronic Markets (Berg et al., 2003) but also leads to more pricing anomalies and slower convergence to terminal cash flows in TradeSports markets (Tetlock, 2008). Experimental results indicate that noise traders lower the informational efficiency of markets (Bloomfield et al., 2006), so the disparate field results may be due to differences in the composition of traders that emerge as volume increases.

A second set of factors affecting information aggregation concerns the complexity of the information and dividend structures in the market. These issues are amenable to laboratory studies given the difficulty in observing and controlling private information in field settings. Early experimental studies by Plott and Sunder (1988) find convergence and efficiency if simple Arrow-Debreu securities are used that pay a fixed dividend if and only if their associated state occurs, the structure of private information is relatively simple (agents are told which state is *not* true), and there is no aggregate uncertainty (combining all private signals reveals the true state perfectly). This result is replicated for a ten-state environment with less informative private signals (draws from an urn) and aggregate uncertainty by Plott (2000); however, this replication uses approximately ninety subjects whereas the earlier laboratory experiments typically include around twelve or sixteen subjects. Markets with more complicated 'tiered' securities (where dividend payments are state-dependent and vary in magnitude across trader types) generate mixed results; having some traders know the state of the world perfectly, common knowledge of the dividend structures for all types, market experience, and a small number of tiered securities all facilitate convergence and efficiency (Plott and Sunder, 1982, 1988; Forsythe and Lundholm, 1990; O'Brien and Srivastava, 1991).

From 2001–2003, John Ledyard, Robin Hanson, David Porter, and others worked to implement a prediction market to forecast political and economic instability in the Middle East (see Hanson, 2007 for details). The state space for this application becomes prohibitively large for any reasonable question of interest; if one wants to predict which of eight countries will experience GDP growth next quarter then $2^8 = 256$ separate securities are needed to capture the possibility that the likelihood of growth in each country depends on growth in the others. Unless the number of traders is large then the simple act of equilibrating all 256 markets (even with complete information) seems overwhelming.[3] Thus, Ledyard et al. (2007) test the performance of a prediction market that uses only eight states—effectively ignoring the cross-country correlations—against five other mechanisms that use all 256 states: a call market that allowed for trading of combinatorial and conditional events like 'X and Y' or 'X given Y', three mechanisms that aggregated individual announcements of belief distributions, and the market scoring rule (MSR) developed by Hanson (2003), which is described below. Using groups of six subjects, the MSR performed the best and the eight-state prediction markets the worst, though the deck was (necessarily) stacked against the prediction market because it did not contain a complete set of securities.

Ledyard et al. (2007) is closest in design to the current paper. We compare the prediction market mechanism to three other mechanisms (an iterated poll, the pari-mutuel mechanism, and the MSR) in a relatively simple environment with only two states and a complex environment with $2^3 = 8$ states, each with only *three* traders per group. The latter environment is sufficiently large relative to the number of traders that we expect equilibration to be hindered by market liquidity shortages and subjects' cognitive limitations, but not so large that a simplification of the state space is necessary for the mechanism to operate.

Past studies have examined each of the mechanisms we test in different environments. McKelvey and Page (1990) study an iterated poll where each individual is paid on the accuracy of their own reports instead of the accuracy of the average report. This iterated poll fully aggregates all private information in theory, but falls somewhat short of that target in the laboratory. Chen et al. (2001)

[3]Another concern is market manipulation by traders with an interest in the prediction generated by the market. Hanson et al. (2006) show in an experiment, however, that the accuracy of outside observers who use market prices to make predictions is not affected by the presence of these biased traders; Hanson and Oprea (2009) confirm theoretically that manipulators may play the same role noise traders in Kyle (1985) and will therefore increase market efficiency.

also show how a poll out-performs a repeated call market with Arrow-Debreu securities as well as the information of the best-informed individual.[4] The pari-mutuel mechanism—used widely in horse-race wagering—has similar theoretical properties to the prediction market: information should fully aggregate if trade occurs, but fully rational risk-averse traders should never have an incentive to trade. Plott et al. (2003) find that 'prices' converge to the rational expectations prediction in a simple environment, but a simple model of trading based on private information alone predicts behavior better in more complex settings. In the field, Thaler and Ziembda (1988) show that pari-mutuels do a reasonably good job of predicting horse racing outcomes, though betters tend to over-bet the unlikely ('long-shot') horses.[5] Theoretically, the market scoring rule (MSR) fully aggregates information if traders are risk averse and *not forward-looking*, but does provide some incentives for traders to misrepresent their information early to take advantage of others' incorrect beliefs later (see Chen et al., 2007 and Sami and Nikolova, 2007 for two analyses of this mechanism). To our knowledge, only Ledyard et al. (2007)—who find that the MSR performs the best among their mechanisms—and this paper have tested the MSR in the laboratory.

We formally introduce the environments and mechanisms used in our study in the following section. Section 3 details the experimental design. Results appear in Section 4, followed by analyses of our four observations in 5. We conclude with a discussion in Section 6.

2 Environments and Mechanisms

We consider an information aggregation problem where the state of the world consists of two dimensions. The first dimension represents some unobservable factor whose value impacts the realization in the second dimension. For example, the underlying monetary policy of a central bank (the first dimension) will affect whether or not the bank chooses to raise interest rates each quarter (the second dimension). Monetary policy is not directly observable, but interest rate movements are. In this setting traders in a prediction market can use the bank's past interest rate changes to infer its monetary policy and, in turn, predict upcoming interest rate movements. If a collection of traders have different information about past interest rate movements (and the underlying conditions of the economy at the time of those movements) then a prediction market or other information aggregation mechanism can be used to generate more reliable predictions about the probability of future rate increases.

In the laboratory environment we represent this inference problem by choosing one of two biased coins (the underlying first dimension) and then flipping the chosen coin one time (the second dimension that agents try to predict). The goal of an information aggregation mechanism

[4] Chen et al. (2001) also adjust the aggregation of individual reports into a single posterior to account for subjects' risk aversion, though their adjustment does not significantly improve accuracy.

[5] Camerer (1998) attempts to manipulate actual horse races by placing and canceling large wagers, but the bettors return the odds to the 'correct' values relatively quickly. Thus, the effects of manipulations are short-lived.

is to predict the probability that the flip will land 'heads'. Subjects privately observe sample flips of the chosen coin, try to infer which biased coin was chosen, and then predict the probability that the one 'true' flip will be heads. The goal of the mechanism designer is to combine these individual predictions into one aggregated prediction that incorporates all subjects' private information.[6]

Formally, the unknown true state of the world in our experimental environment is given by $(\theta, \omega) \in \Theta \times \Omega$ where θ (the coin) is drawn according to the distribution $f(\theta)$ and ω (the outcome of the coin flip) is drawn according to the conditional distribution $f(\omega|\theta)$. Each agent $i \in I$ privately observes K_i signals (sample coin flips) of ω, which we denote by $\hat{\omega}^i = (\hat{\omega}^i_1, \ldots, \hat{\omega}^i_{K_i}) \in \Omega^{K_i}$. Each $\hat{\omega}^i_k$ is drawn according to $f(\omega|\theta)$, so signals provide independent, unbiased information about θ that can then be used to predict the *true* value of ω.

Given the signal $\hat{\omega}^i$ and the priors $f(\theta)$ and $f(\omega|\theta)$, agent i forms a posterior belief $q(\theta|\hat{\omega}^i)$ over Θ using Bayes's rule. For simplicity, we denote this posterior on Θ by $q^i(\theta)$. From this, i forms a posterior over Ω given by $p^i(\omega) = \sum_{\theta' \in \Theta} f(\omega|\theta') q^i(\theta')$.

The goal of the mechanism designer is to aggregate the beliefs of the individual agents. The most accurate posterior the designer could hold in this setting would be that which she would form if she had *full information*, meaning she observes every agent's private signal. Letting $\hat{\omega} = (\hat{\omega}^1, \ldots, \hat{\omega}^I)$, we define $q^F(\theta) := q(\theta|\hat{\omega})$, which leads to the *full information posterior* on Ω given by

$$p^F(\omega) = \sum_{\theta' \in \Theta} f(\omega|\theta') q^F(\theta').$$

To evaluate the performance of a given mechanism we compare the belief distribution over Ω implied by behavior in the mechanism to the full information posterior p^F. Abstracting away from the details, we think of mechanisms as producing a sequence of distributions over Ω denoted by $\{h_t\}_{t=0}^T$. Each distribution h_t represents the posterior at time $t \in \{0, \ldots, T\}$ implied by the messages sent by the players up through time t. Thus, h_0 corresponds to the prior and we refer to h_T as the *output distribution* of the mechanism. At any point t we call h_t the *running posterior* at time t. After observing the mechanism, the mechanism designer takes h_T as his posterior over Ω. *Full information aggregation* occurs whenever the mechanism produces an output distribution equal to the full information posterior, or $h_T \equiv p^F$. When Ω is finite we can measure the 'error' of the output distribution, relative to the full information posterior, by the normalized Euclidean norm[7]

$$\|h_T, p^F\|_\rho := |\Omega|^{1/2} \left(\sum_{\omega \in \Omega} |h_T(\omega) - p^F(\omega)|^2 \right)^{1/2}. \tag{1}$$

[6]Our 'sterile' version of the field setting allows us to test the ability of mechanisms to aggregate information in an (essentially) context-free environment. Our results therefore provide a baseline prediction about the relative performance of various mechanisms for use in any related field application.

[7]The normalization by $|\Omega|^{1/2}$ sets the norm of the centroid vector $(1/|\Omega|, \ldots, 1/|\Omega|)$ equal to one regardless of the size of Ω. This allows for casual comparison of distances between spaces of different dimension, though such comparisons should be made very cautiously.

θ	$f(\theta)$	$f(H\|\theta)$	$f(T\|\theta)$
X	1/3	.2	.8
Y	2/3	.4	.6

Table 1: The distribution f for the 2-state experiments.

Our primary measure of the success of a mechanism is the average (or expected) size of this distance.

Environments

In our experiments we compare two environments that vary in the size of the state space and complexity of the information structure. The simpler environment is described above; one of two biased coins are chosen and, upon flipping, the chosen coin either comes up heads or tails. Since there are two flip outcomes we refer to this as the 'two-state' environment. In the more complex environment three biased and correlated coins are randomly ordered and then all three are flipped in the chosen order. There are eight possible outcomes of the flip of three coins, so we refer to this as the 'eight-state' environment.[8] The two environments are described formally below.

Two-State Environment

In the two-state design, $\Theta = \{X, Y\}$ and $\Omega = \{H, T\}$ with $f(\theta)$ and $f(\omega|\theta)$ given in Table 1. The interpretation is that one of two biased coins (X or Y) is to be randomly selected and flipped one time. The X coin is selected with probability 1/3 and comes up heads with probability 0.20. The Y coin is selected with probability 2/3 and comes up heads with probability 0.40. Agents observe neither the chosen coin (θ) nor the outcome of the flip (ω); instead, each agent observes sample flips of the chosen coin ($\hat{\omega}^i \in \Omega^{K_i}$), uses this information to form beliefs over which coin was selected (X or Y), and then forms a probability estimate that the one 'true' coin flip is heads ($p^i(H)$).

Eight-State Environment

In the eight-state design there are three coins, X, Y, and Z, placed in a random order such as YZX or ZYX. The set Θ contains the six possible orderings, each of which is equally likely *a priori*. Once an ordering is chosen, the three coins are then flipped in that order. The result is a triple of heads and tails, such as HHT or THT, where the first character corresponds to the flip of the first coin in the order, the second character corresponds the second coin, and so on. The set Ω contains all eight possible flip outcomes. Agents do not know the true outcome of the flip of the three coins (ω) nor the actual ordering of the coins (θ); instead, they observe sample flips of the chosen coin ordering

[8]Technically these names are misnomers since the true state spaces ($\Theta \times \Omega$) are of size $2 \times 2 = 4$ and $6 \times 8 = 48$, respectively.

θ	$f(\theta)$	TTT	TTH	THT	THH	HTT	HTH	HHT	HHH
XYZ	1/6	.320	$.21\bar{3}$	.160	$.10\bar{6}$	.040	$.02\bar{6}$	.080	$.05\bar{3}$
XZY	1/6	.320	.160	$.21\bar{3}$	$.10\bar{6}$	.040	.080	$.02\bar{6}$	$.05\bar{3}$
YXZ	1/6	.320	$.21\bar{3}$	.040	$.02\bar{6}$	.160	$.10\bar{6}$	.080	$.05\bar{3}$
YZX	1/6	.320	.040	$.21\bar{3}$	$.02\bar{6}$	.160	.080	$.10\bar{6}$	$.05\bar{3}$
ZXY	1/6	.320	.160	.040	.080	$.21\bar{3}$	$.10\bar{6}$	$.02\bar{6}$	$.05\bar{3}$
ZYX	1/6	.320	.040	.160	.080	$.21\bar{3}$	$.02\bar{6}$	$.10\bar{6}$	$.05\bar{3}$

Table 2: The distribution f for the 8-state experiments.

($\hat{\omega}^i \in \Omega^{K_i}$), use this information to form beliefs over which of the six orderings was selected, and then form beliefs over the eight possible outcomes of the 'true' coin flips ($p^i(HHT)$, $p^i(THT)$, etc.).

The X coin lands heads with probability 0.20 and the Z coin lands heads with probability 0.40. The Y coin is different; its flip matches the flip of the X coin with probability 2/3 and differs from X with probability 1/3. The values of $f(\theta)$ and $f(\omega|\theta)$ for this environment are given in Table 2.

Note that, unconditionally, the Y coin lands heads with probability 0.40, making it indistinguishable from the Z coin if one ignores the correlation between coins. In other words, an agent trying to infer the ordering of the three coins based on a sample of flips must first identify the X coin by its lower frequency of heads and then distinguish between the Y and Z coins by identifying which is correlated with X. When each agent has a small number of sample flips this inference problem is difficult and the value of each agent's private information is small. This is the sense in which the eight-state environment is considered more complex.

One real-world setting with a similar correlation structure is the conference championship structure used in many professional and collegiate sports. Here, coin X represents the event that Team A beats Team B in the Western conference championship, coin Z represents the event that Team C beats Team D in the Eastern conference championship, and coin Y represents the event that the Western conference champion beats the Eastern conference champion in the final match-up. Clearly coin Y depends on which teams actually advance to the final game; thus, Y will be correlated with the other two coins. If probabilities were elicited for only the three games then this correlation would not be captured; it takes a full set of $2^3 = 8$ probabilities to capture this correlation.

Mechanisms

In any field application a mechanism's performance—and, therefore, agents' payoffs—depends on the realized value of ω. Consequently, even mechanisms that fully aggregate information can perform poorly when an unlikely true state happens to occur. In the controlled laboratory setting one way to reduce this variation is to reward subjects based on the *expected* performance of the mechanism given the true distribution $f(\omega|\theta)$.[9] In our experiments we generate an estimate of

[9]This cannot be done in most field settings since θ is not observed.

$f(\omega|\theta)$ using 500 draws of ω. Letting $\phi(\omega)$ denote the fraction of the 500 draws that equals ω, the empirical distribution ϕ serves as a close approximation to the true distribution $f(\omega|\theta)$.[10] Subjects are then paid based on the expected performance of the mechanism given the distribution $\phi(\omega)$. This is explained in more detail with each mechanism.

In what follows we index the elements of Ω by $s \in \{1, \ldots, S\}$. In the two-state environment $S = 2$ and in the eight-state environment $S = 8$.

Double Auction

The standard prediction market mechanism used widely in field applications is a double auction with a complete set of Arrow-Debreu securities, henceforth referred to as the 'double auction' mechanism. Here, S state-contingent securities (one for each $\omega_s \in \Omega$) are traded in separate markets. Subjects buy and sell each security in a standard computerized double auction format with an open book where all bids and asks are public information. Traders are initially endowed with cash but no assets; those who want to sell an asset do so by selling short and holding negative quantities. At the end of the trading period each asset s is worth $\phi(\omega_s)$ experimental dollars. Traders who own a positive quantity of asset s receive $\phi(\omega_s)$ experimental dollars per unit and traders who hold a negative quantity of asset s pay $\phi(\omega_s)$ experimental dollars per unit.[11]

In a rational expectations equilibrium the asset prices reveal all private information. Under certain assumptions about preferences these equilibrium prices will equal the full information posterior probabilities.[12] Thus, we set the mechanism output distribution equal to the vector of security prices. If the market prices of traded securities sum to greater than one we adjust them proportionally and set all non-traded security prices to zero. If the market prices sum to less than one and there are untraded securities then we set the price of all untraded securities equal to each other and such that the resulting sum of all security prices is one. If the market prices sum to less than one and there are no untraded securities we adjust all prices upward proportionally to sum to one.

Since this mechanism is zero-sum, however, the no-trade theorem of Milgrom and Stokey (1982) implies that we should not expect any trade in equilibrium with risk-averse agents. Whether or not trade actually occurs and prices equilibrate to the full information posterior, however, depends on the beliefs, preferences, and rationality of the traders.

Pari-mutuel Betting

In pari-mutuel betting traders buy 'tickets' or 'bets' on each of the S possible states. Tickets cost one experimental dollar each and a trader can buy as many tickets of each type as he can afford

[10] We chose to approximate $f(\omega|\theta)$ using $\phi(\omega)$ because the latter is constructed though an actual (computerized) process; we expect that this makes it more understandable to subjects without a statistics background.

[11] In field applications the asset corresponding to the true state is worth one dollar and all other assets are worthless.

[12] See Manski (2006), Wolfers and Zitzewitz (2006), and Gjerstad (2004).

using his cash endowment. During the period the total number of tickets of each type that have been purchased is displayed publicly. At the end of the period these totals are used to calculate the payoff odds for each security. If T_s is the total quantity of state-s tickets purchased then the payoff odds for state s are given by $O_s = (T_s / \sum_m T_m)^{-1}$. Each state-$s$ ticket is then redeemed for $O_s \cdot \phi(\omega_s)$ experimental dollars. In other words, each ticket is worth the payoff odds times the (approximated) true probability that state ω_s occurs.

The total payoff across all tickets and individuals equals the sum of all purchases, making this a zero-sum game. As in the double auction a no-trade theorem applies, so risk-averse agents should not purchase tickets in an equilibrium with common knowledge of rationality. In the presence of noise trading, however, rational traders may have an incentive to participate. It is certainly the case that, once information has fully aggregated, rational, risk-averse agents will purchase tickets to move the inverse of the payoff odds to the (common) posterior probabilities; for this reason we set the mechanism output distribution equal to the inverse of the payoff odds for each state. Whether or not information will actually aggregate, however, is a question for the laboratory.

Iterative Polls

Iterative polls—an incentivized version of the "Delphi method"—are perhaps the simplest and most direct information aggregation mechanism. Subjects are asked to report simultaneously a probability distribution over Ω. The reports are averaged across subjects (by taking the arithmetic mean of the reports for each state) to generate an 'aggregated' report. This aggregated report is shown to all subjects, who are then asked to submit simultaneously a second distribution over Ω. Subjects' second reports will incorporate their own private information plus any information inferred from the average of the first reports. The average of these second reports is displayed, and the process is repeated for a total of five reports. The fifth average report is then taken as the output distribution of the mechanism.

All subjects are all paid identically based on the accuracy of the final report using a logarithmic scoring rule. Specifically, if $h_T(\omega_s)$ is the final average probability report then for each state ω_s each subject i is given $\ln(h_T(\omega_s)) - \ln(1/S)$ tickets. Thus, agents gain state-s tickets if $h_T(\omega_s) > 1/S$ and lose state-s tickets if $h_T(\omega_s) < 1/S$. Once the empirical frequency ϕ is revealed each state-s ticket pays out $\phi(\omega_s)$ dollars. Since all agents receive the same payment the game is not zero-sum and therefore must be subsidized by the mechanism designer.

The logarithmic scoring rule is incentive compatible (Selten, 1998), so any risk-neutral individual acting in isolation would prefer to announce truthfully her beliefs over Ω. In the multiple-player game there exist sequential equilibria in which full information aggregation occurs; thus, we take the final average announcement to be the mechanism's output distribution. One might conjecture that *any* sequential equilibrium should feature full information aggregation since all players have identical incentives, but in fact there exist 'babbling' equilibria in which full information aggregation does not occur.[13] Under risk neutrality the full information aggregation equilibria are Pareto dominant, so the success of the mechanism depends on agents' ability to coordinate on this payoff-dominant outcome.

Market Scoring Rule

In the market scoring rule (MSR), a probability distribution $h_0 = (h_0(\omega_1), \ldots, h_0(\omega_S))$ is publicly displayed at the beginning of each period; in our experiments, $h_0(\omega_s) = 1/S$ for each s. At any given time t during the period, any trader may 'move' the current distribution h_t to a new distribution, h_{t+1}. This is done simply by announcing the new distribution h_{t+1}. When a trader makes such a move he receives (or loses)

$$\ln(h_{t+1}(\omega_s)) - \ln(h_t(\omega_s)) \qquad (2)$$

state-s tickets for each s. Traders are given an initial endowment of tickets and cannot move h_t to some h_{t+1} if such a move would require surrendering more tickets of some state than the trader currently holds. This prevents traders from moving probabilities arbitrarily close to zero since the logarithm becomes infinitely negative for arbitrarily small probabilities.

During the period traders may move the probability distribution as many times as they like, subject to the budget constraint. With each move, they gain and lose tickets appropriately. At the end of the period each state-s ticket is worth $\phi(\omega_s)$ experimental dollars. Since summing equation (2) over all t yields

$$\ln(h_T(\omega_s)) - \ln(h_0(\omega_s)),$$

the total change in ticket holdings depends only on the starting distribution h_0 and the ending distribution h_T (intuitively, each trader is 'buying out' the position of the previous trader) The final cash value of this difference must be subsidized (or collected) by the mechanism designer.

As in the iterated poll this mechanism uses the logarithmic scoring rule which is incentive compatible for any risk neutral individual, meaning players will truthfully reveal their beliefs if they do not expect to make any future moves. Thus, if it is common knowledge that each player's final move is in fact their last then each will fully reveal their beliefs in the final move and information will fully aggregate in the final move of the period.[14] For this reason we take the final move of the period to be the output distribution of the mechanism.

If a player does expect to move again in the future then there may be an incentive to misrepresent one's information so that other players erroneously move the distribution away from the full

[13] In a 'good' equilibrium each player announces truthfully in the first round, all players use the first average report to infer others' information, then all players announce the full information posterior in rounds two through five, ignoring any deviations by others. In a 'babbling' equilibrium all players submit random, meaningless announcements in rounds one through four, ignore others' announcements, and attempt to maximize their payoff in the final round; since no information was conveyed in the first four rounds, the final average report generically will not achieve full information aggregation.

[14] This argument is based on the analysis of Chen et al. (2007); see also Sami and Nikolova (2007).

information posterior and the misrepresenting player can then earn profits by moving it back. In our experiment players can make moves at any time during the five-minute window, so it is not clear whether manipulations will persist through the final move or whether information will fully aggregate at the end of the period. We test for evidence of manipulations in Section 4.

3 Experimental Design

All experiments were run at the California Institute of Technology using undergraduate students recruited via e-mail. Each period lasted 5 minutes and subjects earned an average of approximately thirty dollars per session.

In each period subjects are publicly informed about the distribution f given in Tables 1 and 2, so we take this as the common prior.[15] A coin (or coin ordering) θ is chosen by the computer but not revealed to the subjects. Instead, each subject is privately shown a unique sample of coin flips of the chosen coin. The mechanism is then run and the output distribution is observed. After the period ends traders are told the chosen coin and the distribution $\phi(\omega)$ generated from 500 sample flips of the chosen coin.[16] Subjects' total earnings are then augmented by their payment for the period and the next period begins.

We employ a 4×2 experimental design in which each of the four mechanisms described in Section 2 is run in both the two-state and eight-state environments described in Section 2. Agents participate in groups of three and are matched with the same group for the entire experiment. Each subject group participates in one mechanism for eight periods followed by a different mechanism for eight periods. We use a crossover design in which the ordering of mechanisms for one group is then reversed for another group. Each ordering is run twice for a total of 16 experimental sessions.[17] Table 3 lists the details of each session.

4 Results

The results are organized as follows: First we describe the four ways in which we measure the performance (or failure) of each mechanism. We then show that behavior does not significantly differ across periods and does not depend on whether a mechanism is presented first or second within a given session, allowing us to aggregate results across periods and orderings and directly compare the four mechanisms using our four performance measures.

[15] Technically, the prior is common *information* but not necessarily common knowledge.

[16] All individual signals are independent and independent of the 500 flips used to determine $\phi(\omega)$.

[17] We pair the pari-mutuel with the MSR and the double auction with the poll. This choice is arbitrary; what matters is that for each pairing we run both orderings of that pairing to test for ordering or learning effects.

Session Number	No. of States	No. of Agents	Mechanism 1 (Periods 1–8)	Mechanism 2 (Periods 9–16)
1	2	3	Pari-mutuel	Market Scoring Rule
2	2	3	Pari-mutuel	Market Scoring Rule
3	2	3	Market Scoring Rule	Pari-mutuel
4	2	3	Market Scoring Rule	Pari-mutuel
5	2	3	Double Auction	Iterative Poll
6	2	3	Double Auction	Iterative Poll
7	2	3	Iterative Poll	Double Auction
8	2	3	Iterative Poll	Double Auction
9	8	3	Pari-mutuel	Market Scoring Rule
10	8	3	Pari-mutuel	Market Scoring Rule
11	8	3	Market Scoring Rule	Pari-mutuel
12	8	3	Market Scoring Rule	Pari-mutuel
13	8	3	Double Auction	Iterative Poll
14	8	3	Double Auction	Iterative Poll
15	8	3	Iterative Poll	Double Auction
16	8	3	Iterative Poll	Double Auction

Table 3: The experimental design.

Measures of Performance

Our primary measure of a mechanism's performance is the average normalized Euclidean distance between the mechanism's output distribution h_T and the full information posterior p^F (see equation (1) above); this provides a simple measure of how accurate the mechanism designer's posterior beliefs are relative to the ideal case of full information aggregation.[18]

One might also be concerned with other properties of the mechanism's performance. For example, consider the no-trade theorem in the context of the prediction market and pari-mutuel mechanisms. In a thin market devoid of noise traders, (weakly) risk-averse rational traders should (weakly) prefer not to participate in either mechanism. If no trade occurs then the mechanism provides no value to the designer since no new information is revealed. If the market were sufficiently thick then it becomes more likely that noise traders will exist—or at least that rational traders believe that noise traders exist—and so trade will occur and information will be revealed. In our experiments, however, groups contain only three agents so the logic of the no-trade theorem is particularly compelling in this setting.

Worse than the no-trade outcome is a situation where the mechanism output is misleading. For example, if a mechanism's output distribution in the two-state environment indicates that heads is *less* likely than previously expected when in fact the private information indicates that heads

[18]Other distance measures such as the Kullback and Leibler (1951) information criterion generate qualitatively similar results.

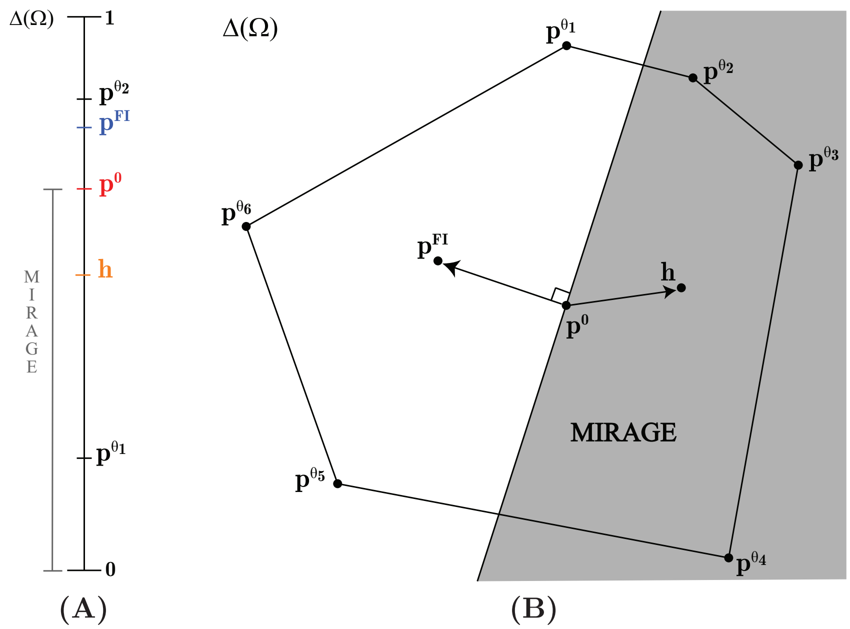


Figure 1: Mirages with (A) two states, and (B) more than two states.

is *more* likely to occur then the designer's posterior is less accurate than the prior. This outcome has been called a *mirage* in the existing literature (Camerer and Weigelt, 1991). In general, we label an output distribution as a mirage if it lies in the opposite direction from the prior as the full-information posterior. Formally, a mirage occurs when $(p^{FI} - p_0) \cdot (h_T - p_0) < 0$, where p^0 is the prior, h_T is the output distribution, and p^{FI} is the full-information posterior. Graphical representations of a mirage (for both two- and eight-state environments) are provided in Figure 1.

A third possible failure of a mechanism is a situation where the output distribution cannot be rationalized by Bayes's rule. We label such outcomes as *Bayes-inconsistent*. For example, the probability of heads in the two-state environment must lie between 0.2 (the probability of heads for the X coin) and 0.4 (the probability of heads for the Y coin).[19] If the mechanism output probability of heads is 0.43 then the logic of standard probability theory offers no advice as to what the best prediction should be; certainly one could construct *ad hoc* theories to rationalize this output and generate a prediction, but from our view this output represents a failure of the mechanism precisely because such *ad hoc* theories become necessary. Graphical representations of Bayes-inconsistent outcomes (for two and eight states) is provided in Figure 2.

For each mechanism in each environment we compare the distance to the full information posterior and count the number of periods in which no trading, mirages, or inconsistencies occur.[20]

[19] For a formal proof of this fact more generally, see Shmaya and Yariv (2007).

[20] We have also constructed various measures of the degree to which each failure occurs; these results are qualitatively similar to counting the number of failures

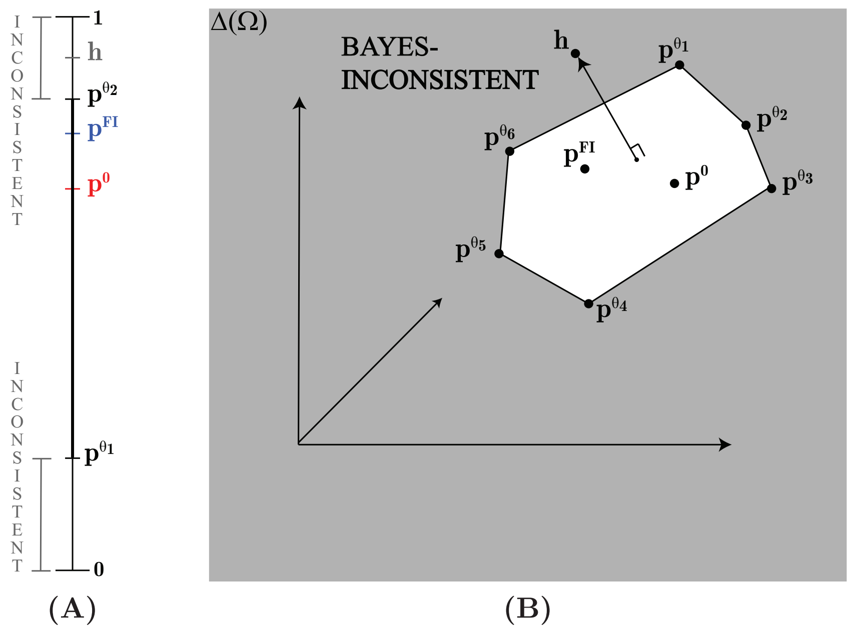


Figure 2: Bayes-inconsistent outcomes with (A) two states, and (B) more than two states.

Period and Order Effects

Although one might expect learning and experience to generate better performance in later periods, we do not find strong evidence for this hypothesis. Using a Wilcoxon rank sum test for equality of medians, we compare the distance between the mechanism output distribution and the full information posterior for each period t against the distance for each period $s \neq t$. Aggregating across all four mechanisms, we cannot reject the hypothesis that the distances have equal medians for any pair of periods in the two-state experiments or in the eight-state experiments. Thus, for example, the distribution of first-period distances has approximately the same median as the distribution of last-period distances, indicating that no significant learning takes place. This is clear from panels (A) and (B) of Figure 3. The same set of tests run on each mechanism (rather than aggregating across all four mechanisms) generates the same results.[21]

Since subjects participate in one mechanism for eight periods and then a second mechanism for a subsequent eight periods, some experience from the first mechanism may spill over into the second mechanism, creating a mechanism ordering effect in our data. Comparing the distance between the mechanism output and the full information posterior for mechanisms run in the first eight periods versus those run in the final eight periods reveals no discernible effect; aggregating across all four mechanisms, Wilcoxon tests reject a significant difference in medians for both the two-state experiments ($p = 0.820$) and the eight-state experiments ($p = 0.850$). The same tests run

[21]Specifically, of the 112 period-versus-period tests, we find that four (or, 3.6% of the tests) are significant at the 5% level in the two-state experiments and none are significant at the 5% level in the eight-state experiments.

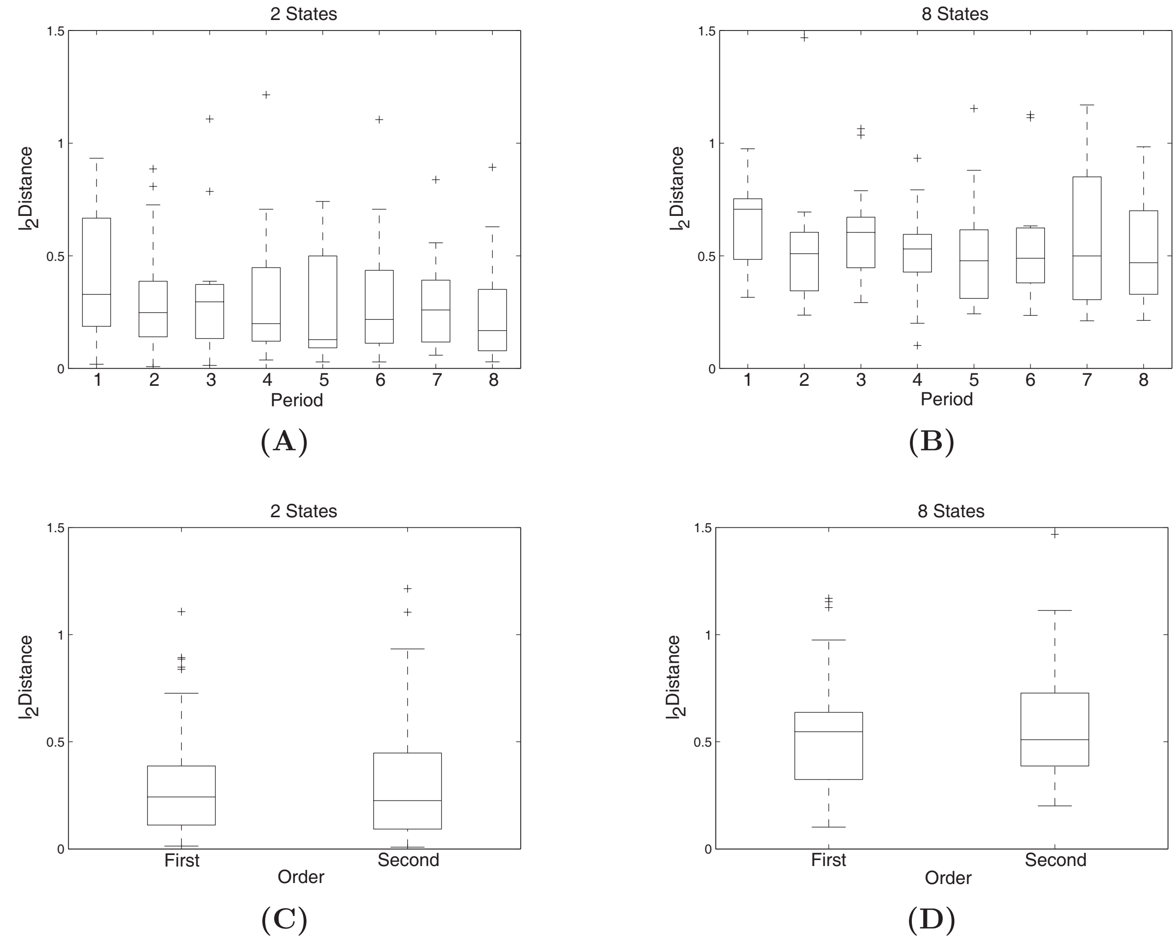


Figure 3: Box-and-whisker plots of the distance between the mechanism output distribution and the full information posterior for (A) each period in the two-state experiments, (B) each period in the eight-state experiments, (C) each mechanism ordering in the two-state experiments, and (D) each mechanism ordering in the eight-state experiments.

on each mechanism individually also find no significant effect (all p-values are greater than 0.168). The plots in panels (C) and (D) of Figure 3 demonstrate this result.

Since we find no significant period or ordering effects, we aggregate across all periods and both orderings in all subsequent analyses.

The Simple Environment: Two States

Mechanism Accuracy

To determine which mechanisms are the most accurate, we perform a comparison of the mechanism error (distance from the mechanism output to the full information posterior) between each pair of

2 States	Avg. Distance	Dbl Auction	Mkt Scoring Rule	Pari-mutuel	Poll
Avg. Distance	-	0.262	0.419	0.295	0.266
Dbl Auction	0.262	-	*0.092*	0.646	0.663
Mkt Scoring Rule	0.419	-	-	0.225	*0.098*
Pari-mutuel	0.295	-	-	-	0.519
Poll	0.266	-	-	-	-
10% Significance Ordering:		MSR $\succeq$ Pari $\succeq$ Poll $\succeq$ DblAuc and MSR $\succ$ Poll $\succeq$ DblAuc			

Table 4: p-values of mechanism-by-mechanism Wilcoxon tests on the distance to the full information posterior for the two-state experiments. Italicized entries are significant at the 10% level.

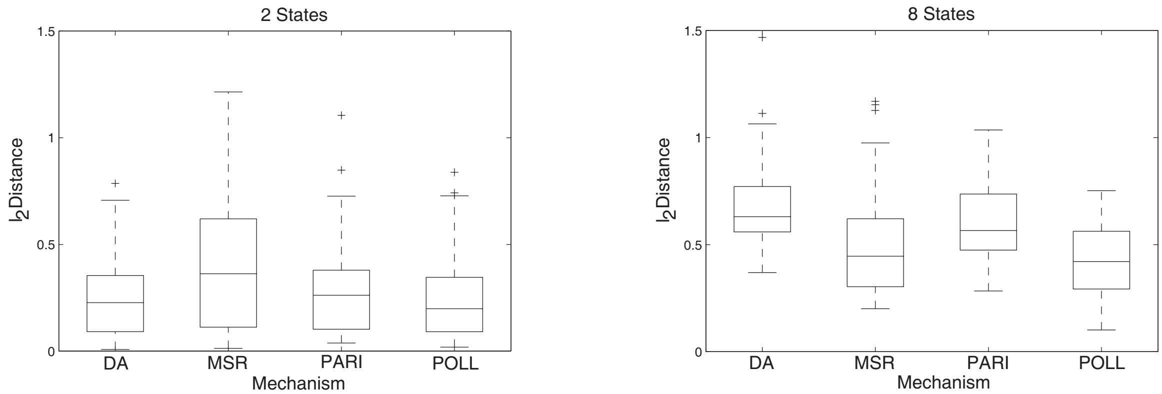
mechanisms. For every given pair, we aggregate across all periods and orderings from the two-state experiments and perform a Wilcoxon test on the resulting distributions of errors. From these comparisons we can construct a 'significance relation' that ranks the four mechanisms according to the degree of error they generate.

Formally, we define the significance relation by $A \succeq B$ if mechanism A has a higher average error than B and $A \succ B$ if that difference is statistically significant at the 10% level. Since $\succ$ is not negatively transitive (it is possible to have $A \not\succ B$ and $B \not\succ C$ but $A \succ C$), describing the relation between mechanisms may require multiple statements. For example, from the pair of statements $A \succeq B \succeq C \succeq D$ and $A \succ C \succeq D$ we conclude that A has significantly higher average error than C and D, but that A's average error is not significantly greater than B's and that no other comparisons are statistically significant.

The result of the pairwise comparison procedure is reported Table 4 and the distributions of errors for each mechanism are shown in panel (A) of Figure 4. The average error for each mechanism is reported in the second row and second column of the table; on average the MSR generates the largest errors and the double auction generates the smallest errors. The p-values of the pairwise Wilcoxon tests are reported in columns three through five and rows three through five. No differences are significant at the 5% level, but the Market Scoring Rule generates significantly higher error than both the poll and the double auction at the 10% level. From this, we generate the significance statements: 'MSR $\succeq$ Pari $\succeq$ Poll $\succeq$ DblAuc and MSR $\succ$ Poll $\succeq$ DblAuc'. Thus, the MSR is the only mechanism that generates significantly higher error than any other mechanism. In other words, these results are not particularly conclusive about which mechanism is the best (in terms of error), but the results are clear about which mechanism is the worst.

Catastrophes: No Trade

In theory, we predict no trade (or indifference to trade) in the double auction and pari-mutuel mechanisms when agents are (weakly) risk averse. In practice (see the second row Table 5), we



(A) **(B)**

Figure 4: Box-and-whisker plots of the distance between the mechanism output distribution and the full information posterior for each mechanism in (A) the two-state experiments, and (B) the eight-state experiments.

2 States	Dbl Auction	Mkt Scoring Rule	Pari-mutuel	Poll
No Trade	0	1	4	0
Bayes-Inconsistent	5	7	6	11
Mirage	13	14	10	12
Bayes-Inconsistent Mirage	0	1	1	3
None	14	12	13	12

Table 5: Number of periods (out of 32) in which each type of catastrophic failure occurs per mechanism in the two-state experiments.

observe trade in each of the 32 periods of the double auction, but no trade in four of the 32 periods (12.5%) of the pari-mutuel mechanism. Despite the fact that it is subsidized, thus circumventing the no-trade issue in theory, we do observe one period of no trade in the MSR. Using a simple binomial test (which assumes independence of no-trade periods) as a rough guide, we conclude that the pari-mutuel mechanism generates no-trade outcomes significantly more frequently than the other three mechanisms (with p-values of 0.008, 0.033, and 0.008 for the double auction, MSR, and poll, respectively). From this we generate the significance statement 'Pari $\succ$ MSR $\succeq$ DblAuc $=$ Poll', indicating that the pari-mutuel is uniquely the worst mechanism in this setting.

Intuitively, we conjecture that subjects are prone to trade, whether rational or not, in the more familiar double auction mechanism and are prone to confusion and, consequently, inactivity in the unfamiliar and mathematically complex Market Scoring Rule mechanism. As for the pari-mutuel mechanism, debriefing discussions with subjects indicated that several believed that first movers would be disadvantaged in this zero-sum game since placing a wager may reveal valuable private information, allowing competitors to gain at the first mover's expense.[22]

Catastrophes: Mirages

The frequency of mirages for the two-state experiments is reported in the fourth row of Table 5. Although all four mechanisms generate a substantial frequency of mirages (ranging from 31% to 44%), the differences between mechanisms are largely insignificant except for marginal significance ($p = 0.0549$) when comparing the MSR to the pari-mutuel.[23] Our significance statement for mirages is 'MSR $\succeq$ DblAuc $\succeq$ Poll $\succeq$ Pari and MSR $\succ$ Pari', indicating that the Market Scoring is uniquely the worst in this setting.[24]

Catastrophes: Inconsistencies

The third row of Table 5 displays the number of periods in which Bayes-inconsistent outcomes occur in the two-state experiments.[25] Clearly the poll is the most frequent; using a simple binomial test we conclude that the poll generates Bayes-inconsistent outcomes significantly more frequently than any of the other three mechanisms (with p-values of 0.013, 0.048, and 0.026 for the double auction, MSR, and pari-mutuel, respectively). Thus, our significance statement regarding Bayes-inconsistency is 'Poll $\succ$ MSR $\succeq$ Pari $\succeq$ DblAuc'. Conditional on observing a Bayes-inconsistent outcome, the average distance between h and the convex hull is 0.024, 0.171, 0.106, and 0.052 for the double auction, MSR, pari-mutuel, and poll, respectively. Thus, the 'magnitude' of the Bayes-inconsistency in the poll is less than in the MSR or pari-mutuel, though it is not clear that this measure is relevant since *all* Bayes-inconsistent outcomes lead to an inference failure, despite the magnitude.

Summary

In each of our four measures (error, no trade, mirages, and Bayes-inconsistencies) we found one mechanism to be uniquely bad and the others to be roughly equivalent. Specifically, the MSR generates the most error, the pari-mutuel generates the most no-trade periods, the poll is the most frequently Bayes-inconsistent, and the MSR creates mirages most frequently. The only mechanism that performed well in all measures (or, did not perform poorly in any one measure) is the double auction mechanism.

[22] In several periods we do observe 'meaningless' trade where a trader submits a wager in the final second before the market closes. If an individual is the only trader to place a wager in a pari-mutuel mechanism, he faces no risk as long as he owns at least one of each security since he is effectively betting against himself. Thus, these trades are not informative (or financially consequential) and are discarded from the analysis.

[23] The number of mirages which are simultaneously Bayes-inconsistent outcomes is 0, 1, 1, and 3 for the double auction, MSR, pari-mutuel, and poll, respectively.

[24] The binomial test between the MSR (with 14 mirages) and the pari-mutuel generates a p-value of 0.0549.

[25] We do find that, across all mechanisms, Bayes-inconsistent outcomes are significantly more likely to occur in the first period. No other period effects have been observed.

8 States	Avg. Distance	Dbl Auction	Mkt Scoring Rule	Pari-mutuel	Poll
Avg. Distance	-	0.696	0.527	0.605	0.418
Dbl Auction	0.696	-	**0.002**	*0.093*	<**0.001**
Mkt Scoring Rule	0.527	-	-	*0.083*	0.324
Pari-mutuel	0.605	-	-	-	**0.001**
Poll	0.418	-	-	-	-
10% Significance Ordering:		DblAuc $\succ$ Pari $\succ$ MSR $\succeq$ Poll			

Table 6: p-values of mechanism-by-mechanism Wilcoxon tests on the distance to the full information posterior for the eight-state experiments. Italicized (bold-faced) entries are significant at the 10% (5%) level.

A summary of the results appears in columns two through five of Table 11.

The Complex Environment: Eight States

Mechanism Accuracy

As with the two-state experiments, we measure a mechanism's error as the l_2 distance between the mechanism output distribution and the full information posterior. The distribution of errors for each mechanism is compared against that of each other mechanism using a Wilcoxon rank sum test. This pairwise comparison procedure generates a significance ordering that ranks the mechanisms by their average errors. The result of this procedure is reported in Table 6. The accuracy results for the eight-state experiments can be summarized by the significance statement 'DblAuc $\succ$ Pari $\succ$ MSR $\succeq$ Poll', which indicates that the double auction is uniquely the worst mechanism (according to this error measure), the pari-mutuel is uniquely the second-worst, and the MSR and poll generate the lowest errors on average, with no significant difference between them.

Catastrophes: No Trade

In the eight-state experiments no-trade periods were observed only in the pari-mutuel mechanism. One group of subjects traded in none of the eight periods and another group failed to trade in only their fifth period. Thus, the pari-mutuel mechanism is uniquely the worst when ranked by the frequency of no-trade periods.

Catastrophes: Mirages

Recall that we define a mirage to be a mechanism output distribution that lies in an opposite direction from the prior as the full information posterior, or when the dot product between $(h - p^0)$ and $(p^{FI} - p^0)$ is negative. This is demonstrated in panel (B) of Figure 1.

8 States	Dbl Auction	Mkt Scoring Rule	Pari-mutuel	Poll
No Trade	0	0	9	0
Mirage	13	7	7	3
None	19	25	16	29

Table 7: Number of periods (out of 32) in which each type of catastrophic failure occurs per mechanism in the eight-state experiments. Note that Bayes-inconsistent is omitted since it occurs in all periods of all mechanisms.

8 States	Avg Angle	Dbl Auction	Mkt Scoring Rule	Pari-mutuel	Poll
Average Angle	-	0.014	0.380	0.246	0.258
Dbl Auction	0.014	-	**<0.001**	**0.011**	**<0.001**
Mkt Scoring Rule	0.380	-	-	0.180	0.773
Pari-mutuel	0.246	-	-	-	0.286
Poll	0.258	-	-	-	-
10% Significance Ordering:		MSR $\succeq$ Poll $\succeq$ Pari $\succ$ DblAuc			

Table 8: p-values of mechanism-by-mechanism Wilcoxon tests comparing the angle between the mechanism output $(h - p^0)$ and the full information posterior $(p^{FI} - p^0)$.

Looking at the frequency of mirages (see Table 7), the double auction is most prone to mirage outcomes while the poll is the least prone. Comparing the distribution of these dot products (which represent the 'angle' between the vectors) and applying pairwise Wilcoxon tests (see Table 8), we see that the double auction is uniquely the worst mechanism in terms of the degree of mirages. A third way to measure the incidence of mirages is simply to count the number of dimensions of $(h - p^0)$ that have the same sign as the corresponding dimension of $(p^{FI} - p^0)$, excluding the first and last dimension since, in theory, they should not change. Table 9 reports the p-values of the pairwise Wilcoxon tests on the number of dimensions. The results are in line with the other measures; the double auction is uniquely the most prone to mirages and the other three mechanisms do not significantly differ in the frequency or magnitude of observed mirages.

Catastrophes: Bayes-Inconsistency

Recall that an output distribution is labeled 'Bayes-inconsistent' if it does not lie in the convex hull of the limit posteriors. In the eight-state case, distributions live in $\mathbb{R}^8$ but since the first and last dimensions should never differ from the prior, the convex hull lives in the six-dimensional subspace where those two dimensions are fixed at the prior level. Thus, an output distribution is automatically 'Bayes-inconsistent' if either the first or last dimension differs from the prior. See Figure 2 for a simplified representation of this issue. In practice, Bayes-inconsistency occurs in every period under every mechanism in our eight-state experiments, so indicating Bayes-inconsistency with a binary indicator variable is not informative. Although any Bayes-inconsistency leads to

8 States	Avg No.	Dbl Auction	Mkt Scoring Rule	Pari-mutuel	Poll
Average No. Dim.	-	2.69	3.69	3.70	3.97
Dbl Auction	2.69	-	**0.002**	**0.003**	**<0.001**
Mkt Scoring Rule	3.69	-	-	0.798	0.239
Pari-mutuel	3.70	-	-	-	0.467
Poll	3.97	-	-	-	-
10% Significance Ordering:			Poll $\succeq$ Pari $\succeq$ MSR $\succ$ DblAuc		

Table 9: p-values of mechanism-by-mechanism Wilcoxon tests comparing the number of dimensions (out of 6) of the mechanism output that move in the same direction (from the prior) as the full information posterior.

difficulties in interpretation, we proceed by measuring the distance between the output distribution and the convex hull. Using pairwise Wilcoxon tests (see table 10), we find that neither the MSR nor the poll have significantly greater median distances than any other mechanism, and that the double auction and pari-mutuel do have significantly greater median distances than at least one other mechanism. Thus, the MSR and the poll are less prone to large deviations from the convex hull.

An alternative way to measure the propensity for Bayes-inconsistency is the count the number of periods in which the distance between the output distribution and the convex hull is within ϵ for each ϵ greater than zero. The resulting graph of frequencies versus ϵ for each mechanism appears in Figure 5. The MSR and the poll generate output distributions within ϵ of the convex hull most frequently when ϵ is small. As ϵ is increased, however, the MSR moves from most frequent to least frequent and the Parimutuel moves from second-least frequent to most frequent. In other words, the MSR output tends to lie either very close to the convex hull or very far, while the pari-mutuel output consistently lies an intermediate distance from the convex hull. Thus, a market observer who is concerned about extreme levels of Bayes-inconsistency should prefer the pari-mutuel mechanism over the MSR in the eight-state environment. As for the double auction mechanism, however, the results are poor in either measure; its average distance from the convex hull is the highest and the frequency with which it lands within ϵ of the convex hull is typically the lowest or second-lowest among the four mechanisms.

Summary

As with the two state case, we found one or two mechanisms to be uniquely bad according to each of our four measures (error, no trade, mirages, and Bayes-inconsistency), though the poorly-performing mechanism varies with the measure. Specifically, the double auction and pari-mutuel generate larger errors, the pari-mutuel is the most prone to no trade, the double auction creates the most mirages, and the double auction and pari-mutuel generate the greatest amount of Bayes-

8 States	Avg Dist	Dbl Auction	Mkt Scoring Rule	Pari-mutuel	Poll
Average Distance	-	0.447	0.362	0.398	0.312
Dbl Auction	0.447	-	**0.001**	0.107	<**0.001**
Mkt Scoring Rule	0.362	-	-	0.180	0.257
Pari-mutuel	0.398	-	-	-	**0.008**
Poll	0.312	-	-	-	-
10% Significance Ordering:			DblAuc $\succeq$ Pari $\succeq$ MSR $\succeq$ Poll		
			DblAuc $\succ$ MSR $\succeq$ Poll		
			DblAuc $\succeq$ Pari $\succ$ Poll		

Table 10: p-values of mechanism-by-mechanism Wilcoxon tests comparing the severity of Bayes-inconsistency, as measured by the distance between the mechanism output distribution and the convex hull of the limit posteriors.

	2 States				8 States			
Summary	Error	No Trade	Mirage	Inconsistent	Error	No Trade	Mirage	Inconsistent
Dbl Auction	✓	✓	✓	✓	×	✓	×	×
M.S.R.	×	✓	×	✓	✓	✓	✓	✓
Pari-mutuel	✓	×	✓	✓	×	×	✓	×
Poll	✓	✓	✓	×	✓	✓	✓	✓

Table 11: Summary of results. A ✓ indicates the mechanism was not significantly out-performed by some other mechanism in that measure and a × indicates that it was.

inconsistency. The one mechanism that did not perform poorly in any of the four measures is the poll. The results for the eight-state experiments are summarized in the last four columns of Table 11.

5 Four Observations

The results indicate that the poll performs well and the double auction poorly in the more complex environment. This raises the deeper question of *why* this occurs; what features of the poll make it successful that are not shared by the double auction? Based on our analysis of the data we state four observations about these two mechanisms that we believe are primarily responsible for the performance differences.

Observation 1 *Preferences are aligned in the poll, so traders have no incentive to misrepresent their information.*

A subject misrepresents his private information when he takes an action intended to send a false signal of his private information. Misrepresentation can interfere with the performance of a mechanism by adding noise to the public signals sent by a subjects actions. While the potential

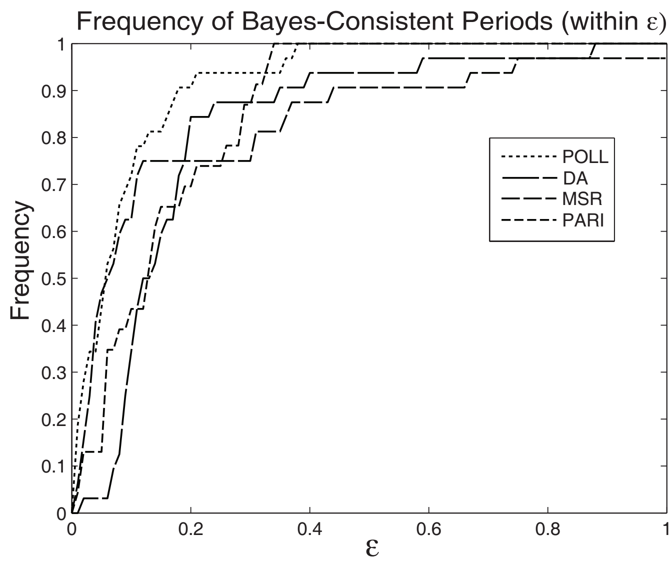


Figure 5: Frequency of periods (with trade) in which Bayes-inconsistency is less than ϵ.

for misrepresentation in equilibrium is a difficult question, it is clear that misrepresentation might present profit opportunities in mechanisms where incentives are not aligned. In the poll, however, a subject's payoff will generally increase in the quality of the information available to other subjects. Thus, the poll may be less subject to problems with misrepresentation.

Misrepresentation in a mechanism should be suspected when actions early in the period move the posterior distribution implied by actions up to that point further away from the full information posterior, while actions late in the period move the implied posterior towards the full information posterior. We construct a rough measure of misrepresentation as follows. Recall that each mechanism generates a sequence of distributions $\{h_t\}_{t=0}^{T}$. An action at time t is said to move the posterior toward the full information posterior if $\|h_t - p^F\| \leq \|h_{t-1} - p^F\|$; otherwise, the action moves the posterior away.[26] A subject is identified as a misrepresenter in a period if his moves include at least one move toward the full information posterior and at least one move away, and all moves away precede all moves toward. We have 96 opportunities to observe misrepresentation for each mechanism (3 subjects each in a total of 32 periods). The number of misrepresentations in each mechanism are presented in Table 12. We observe the fewest instances of misrepresentation in the poll and MSR. This is consistent both with the aligned incentives in the poll and the weak incentive compatibility built into the MSR.

Observation 2 *Traders have an incentive to participate in the poll since it is subsidized.*

[26]For the poll, actions are ordinal and we adopt the convention that $t \in \{0, 2, 4, 6, 8\}$ represent individual reports and $t \in \{1, 3, 5, 7, 9\}$ represent aggregate reports. h_t is then not unique when for t even. The different timing structure for the poll makes formal statistical comparisons difficult.

Mechanism	No. of Misrepresentations
Dbl Auction	14
MSR	5
Parimutuel	12
Poll	3

Table 12: Number of observations of misrepresentation by mechanism

Excluding the value of initial endowments, the double auction is a zero-sum game. A trader who does not participate earns her expected value of participating and, according to the no-trade theorem argument, she strictly prefers non-participation if rationality is common knowledge and she is risk averse. Even if rationality is not known only those traders who expect to perform better than average will prefer to participate. Although trade occurs in every period in our data, there are four periods (of sixty-four) where one of the three traders abstains from trading.

In the poll, however, there is no benefit to abstention; any trader can (weakly) improve the group's average report (relative to his posterior beliefs) by appropriately incorporating his private information into his own final report. Improving the final average report improves the payoff of everyone in the group.[27]

Observation 3 *Traders in the poll must submit entire probability distributions, preventing them from focusing on a small number of securities.*

It appears that market thinness in the eight-state world prevents the double auction from aggregating information properly. We find that there are only 2.60 transactions per minute across all markets in the eight-state environment, compared to 5.00 transactions per minute in the two-state environment; traders are trading half as frequently in the eight-state environment despite the fact that there are four times as many markets. Interestingly, total volume per minute is much higher in the eight-state environment (14.47 units per minute compared to 6.48 units per minute in the two-state environment), indicating that traders in the eight-state environment are making a small number of large transactions. Trades in the eight-state environment tend to focus on a small number of securities. Averaging across the four double auction sessions, trade on the two most active securities accounted for 46 percent of the transactions while trade on the two least active securities accounted for only 8 percent of the transactions.[28]

We conjecture that subjects focusing on a small number of securities indicates that attention is a constraint which binds in mechanisms that require separate focus on each event or security. In the double auction subjects must analyze the market for each security separately. Given bounded

[27]The average payoff per trader per period in the poll is 25.9 cents and 35.0 cents for the two-state and eight-state treatments, respectively.

[28]There does not appear to be a systematic trend in which securities were traded the most or least frequently.

	2 States		8 States	
Mechanism	Last Report	Output Distribution	Last Report	Output Distribution
Dbl Auction	11	11	24	24
Mkt Scoring Rule	18	18	9	9
Pari-mutuel	11	11	9	9
Poll	28	8	21	8

Table 13: Number of periods with far-off last report and final prediction.

attention, subjects are likely to focus or coordinate on a small number of securities, forgoing profits on other securities. Thus, we should expect some market prices to be far from equilibrated. To examine this conjecture we consider the states TTT and HHH, whose posterior probabilities equal the prior probabilities of 24/75 and 4/75, respectively, because the ordering of the coins obviously does not affect the probability of these two states. If market prices are far from these values then profit opportunities may exist in these markets. In fact we observe that the average distance between the final price and the prior probability is 13 percent for TTT and 7.6 percent for HHH. Both of these distances are significantly greater than the distances for any other mechanism (Wilcoxon p-values of < 0.001).

Observation 4 *The poll averages the elicited beliefs, so the effects of a single aberrant trader are mitigated.*

Our final observation is that the poll performs relatively well compared to the other mechanisms due to lessened sensitivity to erroneous last actions. To identify the frequency of large errors in individual reports, we first calculate the average error in final predictions across all the mechanism. Using the normalized l_ρ, the size of average errors in the 2 state experiments is 0.155; in the 8 state experiments it is 0.5996. We define a period with far-off last report as one where the last action implies an individual posterior with a larger-than average prediction error. As the poll requires all three individuals to submit their report simultaneously, we use the report with the largest prediction error as the last report. The number of periods with far-off reports in each mechanism are presented in Table 13.

In the double auction, parimutuel, and MSR, this last report has a direct effect on the mechanism's output. The number of periods with far-off prediction, defined as a period that produces a final prediction with larger-than average prediction error, is necessarily the same as the number of periods with far-off reports. However, the poll balances out this errant last report by averaging it with the other two players' reports.

Despite the large numbers of far-off last reports, the poll produces the fewest instances of far-off final predictions. This is consistent with our claim that averaging in the poll makes it less sensitive to individual errors at end of the period. Note that if the players' final reports are derived from the same distribution, Jensen's inequality and the convexity of our error measure will imply lower prediction error for the poll. Another interesting observation from Table 13 is that the number of far-off last reports in the poll is among the highest compared to other mechanisms, which point to the possibility of players strategically using the averaging mechanism to offset expected error in other players' reports.

6 Discussion

In comparing these four mechanisms (the double auction, the market scoring rule, the pari-mutuel, and the poll), we find that the performance of the mechanisms is significantly affected by the complexity of the environment. In particular, the double auction mechanism appears to perform relatively better when the number of states is small relative to the number of traders and the inference problem of inverting beliefs back into received signals and then converting aggregated signals into an aggregated belief is relatively easy. When the environment becomes more complicated, both in the number of states and in the difficulty of the inference problem, the performance of the double auction market breaks down and other mechanisms emerge as superior. In particular, the iterative poll is the only mechanism in our experiment that was not outperformed by some other mechanism in any of the four measures of error considered.

Identifying which mechanisms perform well in given environments is only the first step in this research. The most compelling line of inquiry is into the underlying reasons for a mechanism to succeed or fail in a given environment. For example, we observe that the failure of the double auction in the eight-state experiments is due in part to the increased ratio of the number of securities to the number of traders: the 'thin markets' problem. As the number of securities exceeds the number of traders, agents apparently focus their limited attention on a small subset of the securities during the trading period. This creates an additional coordination problem as traders seek to focus their attention on markets in which trading is currently most profitable, perhaps due to the trading volume in that market and the private information of the given trader. If some securities are ignored and receive no trades then information aggregation is necessarily incomplete. In general, it would be useful to explore further the boundaries of our results; at what ratio of securities to traders does 'thinness' become a problem? Does correlation between states cause additional problems for information aggregation, or is the difficulty simply due to the sheer number of securities being traded? Further experimental work would provide answers to these questions.

Our larger goal with this line of research is to help develop the practice of behavioral mechanism design, where behavioral insights inform both the design of mechanisms for the immediate future and the modification of theories that can then be used to find optimal mechanisms for practical applications in the long run.

References

Berg, J., R. Forsythe, F. Nelson, and T. Rietz (2003). Results from a dozen years of election futures markets research. In *Handbook of Experimental Economics Results (forthcoming)*. Amsterdam: Elsevier.

Bloomfield, R., M. O'Hara, and G. Saar (2006). The limits of noise trading: An experimental analysis. Cornell University Working Paper.

Camerer, C. F. (1998). Can asset markets be manipulated? A field experiment with racetrack betting. *Journal of Political Economy 106*(3), 457–481.

Camerer, C. F. and K. Weigelt (1991). Information mirages in experimental asset markets. *Journal of Business 64*(4), 463–493.

Chen, K.-Y., L. R. Fine, and B. A. Huberman (2001). Forecasting uncertain events with small groups. In *Proceedings of the 3rd ACM Conference on Electronic Commerce*, pp. 58–64.

Chen, K.-Y. and C. R. Plott (2002). Information aggregation mechanisms: Concept, design, and implementation for a sales forecasting problem. Caltech Social Science Working Paper #1131.

Chen, Y., D. M. Reeves, D. M. Pennock, R. D. Hanson, L. Fortnow, and R. Gonen (2007). Bluffing and strategic reticence in prediction markets. Yahoo! Research working paper.

Cowgill, B., J. Wolfers, and E. Zitzewitz (2009). Using prediction markets to track information flows: Evidence from google. University of Pennsylvania Working Paper.

DeLong, J. B., A. Shleifer, L. H. Summers, and R. J. Waldmann (1990). Positive feedback investment strategies and destabilizing rational speculation. *Journal of Finance 45*, 379–395.

Forsythe, R. and R. Lundholm (1990, March). Information aggregation in an experimental market. *Econometrica 58*(2), 309–347.

Gjerstad, S. D. (2004). Risk aversion, beliefs, and prediction market equilibrium. University of Arizona Working Paper.

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

Hanson, R. (2007). The policy analysis market: A thwarted experiment in the use of prediction markest for public policy. *Innovations 2*(3), 73–88.

Hanson, R. and R. Oprea (2009). A manipulator can aid prediction market accuracy. *Economica 76*, 304–314.

Hanson, R., R. Oprea, and D. Porter (2006, August). Information aggregation and manipulation in an experimental market. *Journal of Economic Behavior and Organization 60*(4), 449–459.

Hopman, J. W. (2007). Using forecasting markets to manage demand risk. *Intel Technology Journal 11*(2), 127–136.

Kullback, S. and R. A. Leibler (1951). On information and sufficiency. *Annals of Mathematical Statistics 22*, 79–86.

Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica 53*, 1315–1336.

Ledyard, J. O., R. Hanson, and T. Ishikida (2007). An experimental test of combinatorial information markets. California Institute of Technology Working Paper.

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

McKelvey, R. D. and T. Page (1990). Public and private information: An experimental study of information pooling. *Econometrica 58*(6), 1321–1339.

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

O'Brien, J. and S. Srivastava (1991). Dynamic stock markets with multiple assets: An experimental analysis. *Journal of Finance 46*(5), 1811–1838.

Plott, C. R. (2000). Markets as information gathering tools. *Southern Economic Journal 67*, 2–15.

Plott, C. R. and S. Sunder (1982, August). Efficiency of experimental security markets with insider information: An application of rational-expectations models. *Journal of Political Economy 90*(4), 663–698.

Plott, C. R. and S. Sunder (1988, September). Rational expectations and the aggregation of diverse information in laboratory security markets. *Econometrica 56*(5), 1085–1118.

Plott, C. R., J. Wit, and W. C. Yang (2003). Parimutuel betting markets as information aggregation devices: Experimental results. *Economic Theory 22*, 311–351.

Rhode, P. W. and K. S. Strumpf (2004). Historical presidential betting markets. *Journal of Economic Perspectives 18*(2), 127–142.

Rosenbloom, E. S. and W. Notz (2006). Statistical tests of real-money versus play-money prediction markets. *Electronic Markets 16*(1), 63–69.

Sami, R. and E. Nikolova (2007). A strategic model for information markets. In *Proceedings of the 8rd ACM Conference on Electronic Commerce*, San Diego, California, pp. 316–325.

Selten, R. (1998). Axiomatic characterization of the quadratic scoring rule. *Experimental Economics 1*, 43–61.

Servan-Schreiber, E., J. Wolfers, D. M. Pennock, and B. Galebach (2004). Prediction markets: Does money matter? *Electronic Markets 14*(3), 243–251.

Shmaya, E. and L. Yariv (2007). Foundations of bayesian updating. California Institute of Technology working paper.

Tetlock, P. C. (2008). Does liquidity affect securities market efficiency? University of Texas Working Paper.

Thaler, R. H. and W. T. Ziembda (1988). Anomalies: Parimutuel betting markets: Racetracks and lotteries. *Journal of Economic Perspectives 2*, 161–174.

Wolfers, J. and E. Zitzewitz (2004). Prediction markets. *Journal of Economic Perspectives 18*(2), 107–126.

Wolfers, J. and E. Zitzewitz (2006, December). Interpreting prediction market prices as probabilities. NBER Working Paper No. 12200.

Appendix A: Experiment Instructions

General

This is an experiment in the economics of decision-making. The instructions are simple, and if you follow them carefully and make good decisions, you may earn a considerable amount of money. You will be paid in cash at the end of the experiment.

Coin Flipping

The experimenter has two coins, X and Y. When flipped, coin X lands Heads with probability 0.2 and Tails with probability 0.8. Coin Y lands Heads with probability 0.4 and Tails with probability 0.6.

In each round of this experiment, one of the coins is randomly picked. Coin X will be chosen with a 1/3 chance, and Coin Y will be chosen with a 2/3 chance. Each participant will then see some sample flips of the chosen coin. In total we will generate 9 flips of the coins. One of you will see 2 flips, one of you will see 3 flips, and one of you will see 4 flips. For example, if you see 3 flips, your information might look like this:

H T H

In this example, the coin came up "Heads" twice and "Tails" once.

Your flips are your own private information. Do not share them with anyone.

Organization

In this experiment, we will participate in two markets: the "Scoring Rule" market, followed by the "Parimutuel" market. *[THIS STATEMENT VARIED APPROPRIATELY BY GROUP. ONLY TWO OF THE FOLLOWING FOUR SECTIONS WERE THEN SHOWN TO SUBJECT.]*

Pari-mutuel

You will start off each round with 20 francs.

Once you observe your private coin flips, you will have the opportunity to purchase tickets from the experimenter, one for "Heads" and one for "Tails", at a price of 1 franc each. Specifically, there will be two types of tickets: "H" and "T".

The experimenter will keep a running tally of the number of each type of ticket that has been purchased. For example, suppose 18 "H" tickets have been purchased. This number will be denoted by N(H). The experimenter would also track N(T). You will see each of these numbers as they change and the total number of all tickets purchased.

To purchase a ticket, call out "[subject number] [ticket] [quantity]." So, for instance, if you are subject #2, and you wish to buy 3 H tickets, call out "2 H 3." The experimenter will repeat it back to you and type it into the system.

The round will last 5 minutes, during which you may purchase up to 20 tickets total. At the end of the round, we will take the chosen coin and flip it 500 times. We will then calculate the fraction of times that each outcome occurred. We can denote this as p(H) and p(T). For example, if "Heads" comes up in 200 of the 500 flips, then p(H) = 0.40.

Each ticket will then pay out a number of francs according to these two formulas:

Odds of H = (Total number of tickets bought) / N(H)

Payoff of H (in francs) = p(H) x [Odds of H]

So, for instance, if 15 H and 12 T tickets were purchased (which is 27 total,) and if in the 500 flips of the coins, 200 came up H, then each H ticket would be worth (200/500) $\times$ (27 / 15) = (0.4) $\times$ (1.8) = 0.72 francs.

Your total earnings for the period will be the francs earned from the tickets you bought, plus any francs you did not use to buy tickets. Note that unused francs do count towards your payoff, but you can't use them in future periods to buy more tickets.

We will run 8 rounds of this pari-mutuel market. Your earnings will be $1 per 20 francs.

Scoring Rule

In this setting, you will receive tickets that pay off depending on the coin flips. There will be two types of tickets: "Heads" and "Tails". You will start each round with 15 tickets of each type.

There will be a public "probability" for each of the 2 outcomes (Heads and Tails), which we will denote by p(H) and p(T). Both of these probabilities will start off equal to 1/2. Since p(T) must equal 1 - p(H), we will only deal with p(H).

After you observe your private coin flips, you will be able to "move the probability" (p(H)) around. If you would like to move the probability from p to r, for example, then you will receive 15 $\times$ (ln[r] - ln[p]) H tickets. Note that this also moves p(T) from 1-p to 1-r, so you also receive 15 $\times$ (ln[1-r] - ln[1-p]) T tickets.

For example, if the current probability on H is 0.10, and you move H to 0.15, you would receive 15[ln(.15)-ln(.10)] = 6.08 "H" tickets and lose 15[ln(.85)-ln(.90)] = -0.86 "T tickets."

The round will last 5 minutes, during which time anyone may "move the probability." The only limitation is that you may not move the probability if doing so will cause you to run out of one kind of tickets.

At the end of the round, each type of ticket will pay off as follows:

The chosen coin will be flipped by the experimenter 500 times. Each heads ticket will be worth the fraction of heads and each tails ticket will be worth the fraction of tails.

For instance, if 200 of the flips turned out to be Heads, each "H ticket" would be converted into 0.20 francs, etc.

We will run 8 rounds, and francs will be converted into dollars at a rate of $1 per 20 francs.

Poll

In this setting, you will receive tickets that pay off depending on the coin flips. There will be two types of tickets: Heads (H) and Tails (T). You will start each round with 20 tickets of each type.

At the end of each round, we will flip the chosen coin 500 times. Each ticket will pay off an amount equal to the fraction of flips that match the ticket type. If you have a negative holding of a ticket, you must pay out what it is worth. For example, if the coin comes up Heads in 200 of the 500 flips, then each ticket "H" will pay 0.40 francs. If you own 100 H tickets, you would receive 40 francs, for example, and if you own -10 H tickets, then you must pay out 4 francs.

During the round, you will be asked to report to the experimenter your estimate of the frequency with which each of the 2 possible outcomes will occur. In other words, you must announce your estimate of the frequency that "H" and "T" will occur. The experimenter will publicly announce the average of all the reports for each possible outcome (the average announced frequency of "H" and the average announced frequency of "T".) You will then be asked to supply a new report. This process will repeat 5 times.

At the end of the round, we will take the average of the final set of reports and each of you will gain or lose tickets based on these averages. Specifically, if the average report of outcome "H" is denoted p(H), then everyone will receive 50 × (ln[p(H)] - ln[1/2]) H tickets. Note that number will be negative if p(H) ¡ 1/2, which means everyone would lose some H tickets. The formula is the same for the other possible outcome (T).

So, for instance, if the final poll result is (H,T) = (0.7,0.3), you would receive 16.82 additional H tickets, and you would surrender 25.54 T tickets.

We will run 8 rounds, after which francs will be converted into dollars at a rate of $1 per 20 francs.

Double Auction

You will start off each round with 20 francs.

Once you observe your coin flips, we will run a double auction market for "Tickets." There will be two types of tickets, one for "Heads" and one for "Tails". We denote them by H and T.

You will buy and sell these tickets in a market. You may sell tickets you do not own - that is, you may have a negative number of tickets. The only constraint is that if you have a negative number of tickets, you must have at least half that number of francs on hand. For example, if you want to sell "H" tickets until you end up with -8 "H" tickets, you must hold at least 4 francs to do so. We call this your "bankruptcy constraint". Trades that would cause you to violate the bankruptcy constraint will not be allowed.

Trades will be conducted in a program called "jMarkets", which is described on a separate sheet. **[Available upon request.]**

After 5 minutes, the market will close, and the tickets will pay off as follows:

We will take the chosen coin and flip it 500 times. Each ticket will pay off an amount equal to the fraction of flips that match the ticket type. If you have a negative holding of a ticket, you must pay out what it is worth. For example, if the coin comes up Heads in 200 of the 500 flips, then each ticket "H" will pay 0.40 francs. If you own 10 H tickets, you would receive 4 francs, for example, and if you own -10 H tickets, then you must pay out 4 francs.

We will run 8 rounds of this market, after which francs will be converted into dollars at a rate of $1 per 20 francs.