Information Aggregation Mechanisms and Environmental Complexity

Paul J. Healy John O. Ledyard Sera Linardi J. Richard Lowery*

March 23, 2007

Abstract

We experimentally test four different information aggregation mechanisms. Each mechanism is tested in a simple 'two-state' environment and a complex 'eight-state' environment where the number of securities is large (relative to the number of traders) and the Bayesian inference problem is significantly more difficult. We find that the Double Auction market mechanism performs well in the simple environment, but performs poorly in the complex environment. An Iterative Polling mechanism performs the best in the complex environment but exhibits failures in the simple environment. Thus, the optimal mechanism choice depends crucially on the complexity of the environment.

JEL: C72, C91, D82

Keywords: Information aggregation, prediction markets, mechanism design.

DRAFT COPY: DO NOT DISTRIBUTE OR CITE

*Healy: Department of Economics, The Ohio State University, Columbus, OH 43210. Ledyard and Linardi: Division of Humanities and Social Sciences, California Institute of Technology, Pasadena, CA, 91125. Lowery: Tepper School of Business, Carnegie Mellon University, Pittsburgh, PA 15213. Email: healy.52@osu.edu, jledyard@hss.caltech.edu, slinardi@hss.caltech.edu, and jlowery@andrew.cmu.edu.

1 Introduction

Information mechanisms attempt to aggregate diffuse private information in a manner that will be useful to decision makers. This generally takes the form of generating a posterior distribution over some set of outcomes that, ideally, makes full use of the private information of a number of different agents. In recent years theoretical, experimental, and field research has sought to determine the effectiveness of such mechanisms. While evidence along all dimensions is mixed, the basic theme of the literature is that theory suggests certain types of mechanisms cannot work, while experimental evidence indicates that they do work. Furthermore, different information mechanisms exist, and the characteristics of each mechanism are relevant for their performance, both in theory and in practice. In particular, there is ample reason to believe that certain mechanisms tend to fail in potentially seriously misleading ways when conditions are not ideal; when traders are informationally large or the underlying problem is complicated information mechanisms may exhibit a breakdown in trade where no information is revealed, a confused outcome where the results of the mechanism are incompatible with reasonable prior beliefs, or mirages, where the mechanism suggests a reasonable result that is in fact at odds with the aggregate signal available to agents.

This experimental study compares the performance of four information mechanisms that have received study. The classic information mechanism, and one studied here, is the double-auction-style market for securities that pay off when a specified event occurs. The early work on information mechanisms generally focuses on such markets (Forsythe et al., 1982; Plott and Sunder, 1982, 1988). This mechanism has proved successful under certain conditions in achieving a fully informative rational expectations equilibrium. This result holds despite the fact that the market should suffer from a no-trade problem, indicating that theoretical results are at odds with experimental outcomes. Additionally, when a market has few traders relative to the number of securities, it may suffer from a thin markets problem. Trading may not be sufficiently active in enough securities to meaningfully aggregate agents private information.

Also considered is a parimutuel betting system, where individuals purchase tickets associated with an event and, when said event occurs, receive a payoff determined by the ratio of tickets for that event sold to total tickets sold. This mechanism also should suffer from the no-trade problem that must be overcome with a subsidy (Plott et al., 2003). However, ample evidence from real world parimutuels demonstrates not only that trade occurs even in the presence of a negative subsidy in the form of the track take, but also that the odds implied by betting behavior are closely related to the underlying probabilities and robust to attempts at market manipulation (Camerer, 1998). These odds do, however, tend to suffer from the famous favorite-longshot bias (Snowberg and Wolfers, 2006). More troubling, Plott et al. (2003) and Roust and Plott (2005) find evidence that parimutuel prediction markets are prone to mirages; they can send relatively strong signals about the aggregate posterior probabilities that are at odds with the actual posterior that an agent with access to all private information would form.

An iterative poll similar to that used in McKelvey and Page (1990) is also tested. In this poll, subjects are paid based on the quality of the average report in the last period. There is no concern about a no-trade problem. However, McKelvey and Page (1990) find incomplete information aggregation in this environment, so it is not clear that iterative polling is effective in practice.

Finally, a relative newcomer to the information mechanism family is considered. The market scoring rule mechanism, which is the only mechanism designed specifically for the aggregation of diffuse private information (Hanson, 2003), allows each subject to move a publicly observed probability based on his private information. The payoffs are designed such that it is incentive compatible for agents to reveal their private information if agents act in isolation.[1] This mechanism, however, requires a patron; it is not self-funding like a parimutuel or a double auction market. And, the experimental evidence on the market scoring rule to date indicates that it does not perform as well as theory would predict.

Each mechanism is applied to the same information aggregation problems. The goal of the experiments is to determine which mechanisms will be useful in a variety of potential applications where the number of informed traders may be small, individual traders may have large information size, and problems may be more complex than the standard draws from an urn used in much of the literature. To accomplish this, sessions all involve only three subjects. Subjects receive random draws of private information of varying degrees of informativeness. Consequently, traders will be generally informationally large. Furthermore, half of the sessions involve a difficult mapping from private information to underlying states of the world to realized outcomes. This presents challenges to the mechanisms beyond the obvious that individuals may have difficulty interpreting their private signal. All of these mechanisms rely in part on an iterative element where information revealed by other agents' actions is used to update beliefs; when these actions are hard to interpret the assumption that agents update as perfect Bayesians becomes even more unrealistic than usual.

The null hypothesis maintained throughout the experiment is that the degree and quality of information aggregation is invariant in the information mechanism chosen. If instead the idiosyncratic characteristics of each mechanism interact in a consistent manner with the heuristics subjects use to deal with the difficult information problem or the difficulties encountered with informationally large agents, certain mechanisms will perform better than others.

Our results indicate that the double auction market performs best when the information problem is simple, while the iterative poll outperforms other mechanisms when the problem is relatively complicated.

[1] If players anticipate that others will observe and react to their announcements, then they may have an incentive to misrepresent their information early to affect the path of play.

The market scoring rule, despite its theoretical advantages, performs poorly in all treatments, while the parimutuel performs reasonably well when subjects choose to trade but often results in periods with a total breakdown in trade. These results indicate that the appropriate information mechanism for a given problem will depend upon characteristics of the problem; theoretical and experimental work is necessary to further explore the performance of different mechanisms and to understand the causes for success and failure in various scenarios.

2 Notation and Definitions

The set of traders is $\mathcal{I} = \{1, \ldots, I\}$, where $I = 3$ for our experiments. The set of states of the world is $\Theta \times \Omega$, where $\theta \in \Theta$ represents a possible randomizing device (in our case, this is either a single coin or a possible ordering of three coins) and $\omega \in \Omega$ represents a possible outcome of the randomizing device (e.g., heads or tails.) In our '2-coin' experiments, $\Theta = \{X, Y\}$ (coin X and coin Y) and $\Omega = \{H, T\}$ (heads or tails.) In the '3-coin' experiments,

$$\Theta = \{XYZ, XZY, YXZ, YZX, ZXY, ZYX\}$$

(the possible orderings of coins X, Y, and Z,) and

$$\Omega = \{TTT, TTH, THT, THH, HTT, HTH, HHT, HHH\}$$

(the possible outcomes of flipping the three coins, where HHT means the first two coins came up heads and the third came up tails, for example.) From now on, we refer to these as the 2-state and 8-state experiments, respectively, to reflect the size of Ω.

Let $f(\theta, \omega)$ be the joint distribution over all states of the world, let $f(\theta)$ be the marginal distribution on Θ, and let $f(\omega|\theta)$ be the conditional distribution on Ω given $\theta \in \Theta$. In other words, if I know the coin ordering is θ, then I know the probability that the coin flip will be ω is $f(\omega|\theta)$. Tables 1 and 2 completely describe the distribution f used in our experiments for the 2-state and 8-state cases, respectively.

The objects of interest are agents' beliefs over Ω, which are determined indirectly by beliefs over Θ. In general, beliefs over Θ will be denoted with a q, beliefs over Ω will be denoted with a p, and they will always be related according to the formula

$$p(\omega) = \sum_{\theta \in \Theta} q(\theta)\, f(\omega|\theta). \tag{1}$$

Before any information is revealed, agents share a common prior over Θ of q^0 given by $q^0(\theta) = f(\theta)$. Using equation (1), q^0 induces a prior p^0 over Ω.

Subjects see signals that give them information about θ. They then use this information (and the correlation structure in f) to derive a posterior over Ω via Bayes's Rule. Specifically, agent i sees signal

$$s_i = (s_{i1}, \ldots, s_{ik}, \ldots, s_{iK_i}),$$

where each s_{ik} is an element of Ω. Letting θ^A represent the true ('actual') coin ordering, each s_{ik} is independently drawn from Ω according to $f(\cdot|\theta^A)$. In other words, subject i sees K_i independent coin flips from the true coin ordering. Note that K_i may not equal K_j for $i \neq j$, so agents may be differentially informed. Abusing notation slightly, let $f(s_i|\theta)$ be the probability that signal s_i was generated given coin ordering θ. Then, by Bayes's Law, i's posterior over Θ is

$$q(\theta|s_i) = \frac{f(s_i|\theta)\, f(\theta)}{\sum_{\theta' \in \Theta} f(s_i|\theta')\, f(\theta')},$$

which induces a posterior $p(\cdot|s_i)$ over Ω via equation (1). For brevity, we denote these distributions as $q^i(\theta)$ and $p^i(\omega)$, respectively. The distributions p^i are called the **individual posteriors**.

'**Full information**' refers to observing the entire signal $s = (s_1, \ldots, s_i, \ldots, s_I)$. Using the same math as above, we can construct $q(\theta|s)$ and $p(\omega|s)$, which we can denote by q^F and p^F. The distribution p^F is the **full-information posterior**.

If the true coin ordering were known to be θ (which would almost always occur, for example, as K_i approaches infinity,) then the posterior on Ω would be $p^\theta(\omega) = f(\omega|\theta)$. We refer to each p^θ as a **limit posterior** since (with probability one) it is the limit of the sequence of individual posteriors as K_i grows without bound. The distribution $p^A = p^{\theta^A}$ is the **true limit posterior** since it is the posterior if the true coin ordering were known.

Given s, the maximum likelihood value of θ is

$$\theta^{ML}(s) = \arg\max_{\theta \in \Theta}\ f(s|\theta).$$

To shorten notation, we let $\theta_i^{ML} = \theta^{ML}(s_i)$, $p^{ML} = p^{\theta^{ML}(s)}$ and $p^{MLi} = p^{\theta_i^{ML}}$. The distributions p^{ML} and p^{MLi} are the **maximum-likelihood (ML) posterior** and **individual ML posterior**, respectively.

Regardless of s, it must be that the prior p^0, the individual posteriors p^i, the maximum likelihood posteriors p^{ML} and p^{MLi}, and the full-information posterior p^F are all elements of the convex hull of the limit posteriors, denoted by $P = co(\{p^\theta\}_{\theta \in \Theta})$. Since each of these distributions lives in $\mathbb{R}^N$ when there are N states, we can think of them as N-vectors (or $(N-1)$-vectors since they sum to one.) For example, if

$\Omega = \{\omega_1, \ldots, \omega_n, \ldots, \omega_N\}$, then $p^\theta = (p_1^\theta, \ldots, p_N^\theta)$, where $p_n^\theta = p^\theta(\omega_n)$. In our experiment, $P = [.2, .4]$ when $N = 2$.When $N = 8$, note that p_1^θ and p_N^θ are independent of θ, so $P \subseteq \mathbb{R}^5$ (two dimensions are fixed and the remaining six dimensions must sum to a known constant, so five degrees of freedom remain).

We will abstract away from the details of the mechanisms themselves. Instead, we say that a mechanism begins at some prior distribution h^0 over Ω and ends at some distribution h over Ω. The final distribution h is called the **mechanism output distribution**, or simply **output distribution**. To discuss dynamics, we say that the mechanism generates a sequence of distributions $\{h^t\}_{t=0}^T$ where $t \in [0, T]$ indexes time, h^0 is the prior distribution and $h^T = h$. **Information aggregation** occurs when the output distribution matches (or when h^t converges toward) the full information posterior ($h = p^F$.) A **mirage** occurs when the output distribution matches (or when h^t converges toward) a limit posterior p^θ that is not 'close to' the full information posterior, although the details of this definition are still vague.

If we want to dive into the details of the mechanism, let X_i^t be i's strategy space at time t and x_i^t be i's chosen strategy. The path of play for i is then $x_i = (x_i^0, x_i^1, \ldots, x_i^T)$. Let $x^t = (x_1^t, \ldots, x_i^t, \ldots, x_I^t)$ and $x = (x^1, \ldots, x^T)$. At each t, h^t is a function of $\{x^s\}_{s=1}^t$, so we should write $h^t(\omega; x^1, \ldots, x^t)$. Payoffs to players are then given by $\pi_i(x_i, x_{-i}; \omega^A)$, which may also be written as $\pi_i(h^0, \ldots, h^T; \omega^A)$, or sometimes just $\pi_i(h; \omega^A)$ when only the final distribution matters.

In order to talk about information aggregation and mirages, we need a notion of distance between the output distribution and some target distribution, such as p^F. The Kullback-Leibler distance between a 'true' distribution p and another distribution h is given by

$$KL(h,p) = \sum_{\omega \in \Omega} p(\omega)\,[\log p(\omega) - \log q(\omega)].$$

The KL distance, introduced by Kullback and Leibler (1951), is based on informational entropy (Shannon, 1948), but is not symmetric and does not satisfy the triangle inequality, so it is not a proper distance metric. As an alternative, one could measure the distance between p and h via the (normalized) l_ρ-norm for any $\rho \in \{1, 2, \ldots\} \cup \{\infty\}$, where

$$l_\rho(h,p) = (|\Omega|^{\rho-1} \sum_{\omega \in \Omega} |p(\omega) - h(\omega)|^\rho)^{1/\rho}.$$ [2]

We normalize by $|\Omega|^{(\rho-1)/\rho}$ (where $|\Omega|$ denotes the number of states in Ω) so that the values of the norm lie roughly on the same scale regardless of the number of states.[3] The ε-neighborhood of a distribution p under l_ρ is given by $\mathcal{N}_\rho^\varepsilon(p) = \{q : l_\rho(q,p) < \varepsilon\}$.

[3] Under this normalization, if each component of $p-h$ is equal to $1/|\Omega|$ then $l_\rho(h,p) = 1$ regardless of $|\Omega|$. This normalization is only for convenience in scaling our data; we will never compare distances between 2-state and 8-state experiments.

Finally, we might assume there is some decision maker who takes the mechanism output and chooses an action $y \in Y$. His payoff is $u(y, \omega^A)$, where ω^A is the 'true' coin flip. For our purposes, ω^A is never observed (or generated,) so we take the expected payoff either using the full-information posterior,

$$U^F(y) = \sum_{\omega \in \Omega} p^F(\omega)\, u(y, \omega),$$

the true limit posterior,

$$U^A(y) = \sum_{\omega \in \Omega} p^A(\omega)\, u(y, \omega),$$

the maximum likelihood limit distribution,

$$U^{ML}(y) = \sum_{\omega \in \Omega} p^{ML}(\omega)\, u(y, \omega),$$

or any other distribution over Ω. Let y^F, y^A, and y^{ML} be the maximizers of U^F, U^A, and U^{ML}, respectively, and note that y^F and y^{ML} are functions of s and y^A is a function of θ^A. The decision maker does not know s or θ^A, and so cannot use y^F, y^A, or y^{ML}. Instead, he chooses y to maximize

$$U(y, h) = \sum_{\omega \in \Omega} h(\omega)\, u(y, \omega),$$

where h is the mechanism output. Let $y^*(h)$ be the maximizer of $U(y, h)$.

Given u, we can calculate the decision maker's ex-ante expected loss, relative to the full-information posterior, to be

$$L^F(h) = \sum_{\theta} \sum_{s} \sum_{\omega} [u(y^F(s), \omega) - u(y^*(h(s)), \omega)]\, f(s|\theta)\, f(\omega|\theta)\, f(\theta).$$

Other loss measures can be similarly constructed using other benchmarks, such as y^A, or using different levels of information, such as interim or ex-post expected losses.

3 The Mechanisms

We experimentally compare four mechanisms in the above environment: The double auction, pari-mutuel betting, iterated polls, and a market scoring rule. In each period, subjects begin with the prior beliefs given by f in Tables 1 and 2. A true coin ordering θ^A is drawn, but not revealed to the subjects. Instead, each subject is privately shown a signal s_i from $f(\cdot|\theta^A)$. The mechanism is then run and the outcome distribution h is observed. After the period ends, traders are told the true state θ^A. Subjects' payments then depend on the true state ω^A. Since ω^A is drawn from $f(\cdot|\theta^A)$, subjects' payments are subject to variation through the variation in the draw of ω^A. To reduce this variation, we draw 500 'true' states, $(\omega^{A,1}, \ldots, \omega^{A,500})$, and calculate the frequency with which each state is drawn:

$$\phi(\omega) = \frac{\#\{r : \omega^{A,r} = \omega\}}{500}.$$

Each subject is then paid

$$\pi(x; \phi) = \sum_{\omega \in \Omega} \phi(\omega)\, \pi(h; \omega).$$

This approximates the expected payment when θ^A is known and ω^A is not:

$$E\pi(h; \theta^A) = \sum_{\omega \in \Omega} f(\omega|\theta^A)\, \pi(h; \omega).$$

3.1 Double Auction

In the double auction mechanism, N state-contingent securities (one for each $\omega_n \in \Omega$) 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 n is worth $\phi(\omega_n)$. Traders who own a positive quantity of asset n receive $\phi(\omega_n)$ experimental dollars per unit, and traders who hold a negative quantity of asset n pay $\phi(\omega_n)$ experimental dollars per unit.

Under a rational expectations equilibrium, asset prices are fully revealing, meaning that the full information posterior can be calculated from the vector of prices. Under certain assumptions about preferences, the prices should in fact equal the full information posterior probabilities, and so we set h_n in our analysis equal to the closing price of the state-n security.

One issue with the double auction, however, is that it is a zero-sum game in which the agreeing-to-disagree theorem of Aumann applies, so we should not expect trade in equilibrium with risk averse agents. Whether trade occurs and whether prices equilibrate to the full information posterior are questions to be addressed in the laboratory.

3.2 Pari-mutuel Betting

In pari-mutuel betting, traders buy 'tickets' or 'bets' on each of the N possible states. Tickets cost one experimental dollar each and a trader can buy as many tickets of each type as he can afford. During the period, the number of tickets of each type that has been purchased is displayed publicly. At the end of the period, the total number of tickets purchased of each type, denoted by $(T^1, \ldots, T^n)$, is used to calculate the payoff odds for each state. Payoff odds for state ω_n equals the inverse of the proportion of state-n tickets purchased, or

$$O^n = \left(\frac{T^n}{\sum_{m=1}^{N} T^m}\right)^{-1}.$$

If subject i owns T_i^n state-n tickets, then his total payoff is

$$\pi_i(T_i^1, \ldots, T_i^n; \phi) = \sum_{n=1}^{N} T_i^n \, \phi(\omega_n) \, O^n.$$

As in the double auction, this mechanism is a zero-sum game, and so risk-averse agents should not buy tickets. Conditioning on trade, however, the payoff odds on each ω_n should converge to $(p_n^F)^{-1}$, the full-information posterior.[4] Thus, we set $h_n = 1/O^n$.

3.3 Iterative Polls

Under iterative polls, subjects are simply asked to report a probability distribution over Ω. These reports are averaged across subjects and the average report is publicly displayed. Subjects then re-submit a new probability distribution and the new average is displayed. This process repeats five times. Letting, h_n be the average probability report on ω_n in the fifth poll, each subject i is given $T_i^n(h_n)$ state-n 'tickets', where

$$T_i^n(h_n) = \sum_{n=1}^{N} [\ln(h_n) - \ln(1/N)].$$

Note that $T_i^n(h_n)$ is positive if $h_n > 1/N$ and negative if $h_n < 1/N$. Each state-n ticket is then worth $\phi(\omega_n)$. Thus,

$$\pi_i(h; \phi) = \sum_{n=1}^{N} \phi(\omega_n) \, T_i^n(h_n).$$

Under this payoff formula, we conjecture the existence of an equilibrium in which agents fully reveal their information by the final period, leading to full information aggregation, although it remains to verify the details of this equilibrium. Note that all subjects receive the same number of tickets and therefore receive the same total payoff. Since total payoffs can be positive, this is not a zero-sum game and is therefore not subject to the no-trade issues of the double auction and pari-mutuel betting system.

[4] This is only a conjecture at this point; we need to verify this argument.

3.4 Market Scoring Rule

In the market scoring rule, a probability distribution $h^0 = (h_1^0, \ldots, h_N^0)$ is publicly displayed at the beginning of each period. In our experiments, $h_n^0 = 1/N$ for each n. At any given time t during the period, any trader may 'move' the distribution to a new distribution, h^t. If a trader moves the distribution from h^{t-1} to h^t, then he receives (or loses)

$$T_i^n(h^{t-1}, h^t) = \ln(h_n^t) - \ln(h_n^{t-1})$$

state-n tickets for each n. Traders are given an initial endowment of tickets and cannot move h^{t-1} to some h^t 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. With each move, they gain and lose tickets appropriately. At the end of the period, $h = h^T$ and each state-n ticket is worth $\phi(\omega_n)$, so

$$\pi_i(h; \phi) = \sum_{n=1}^{N} \phi(\omega_n)\, T_i^n,$$

where T_i^n is i's final holdings of state-n tickets.

The market scoring rule is incentive compatible when subjects announce their reports in isolation, though the properties of the equilibria of the dynamic multi-player game is still an open question. The mechanism does not induce a zero-sum game, so trade should occur and converge to the full-information posterior if information aggregates properly. Unlike the iterative polls, however, different traders can earn different payoffs.

4 Experimental Design

We employ a 4×2 experimental design in which each of the four mechanisms described in Section 3 is run in both the two-state and eight-state environments described in Section 2. Three agents are used in all eight treatments. 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. Table 3 lists the details of each session.

All experiments were run at the California Institute of Technology using student subjects recruited via E-mail. Each period lasted 5 minutes (or, for the iterative polls, until the five iterations were complete) and subjects earned an average of around thirty dollars per session.

5 Results

The results are organized as follows: First, we verify that period effects and order effects are not present, which then allows us to aggregate data across periods and across mechanism orderings. We then examine the results of the 2-state experiments, looking first at the accuracy of the mechanism output, relative to the full information posterior benchmark. We also examine the frequency and severity of 'catastrophic' failures, such as periods with no trade, mechanism output distributions that cannot be rationalized by Bayes's Rule ('confused' outcomes) and mechanism output distributions that move in a direction away from the full information posterior ('mirage' outcomes). We then repeat these measures for the more complicated 8-state experiments. Finally, we look to individual-level data for further understanding of the successes and failures of each mechanism.

Throughout this paper we report distances using only the (normalized) l_2 norm. Results for other distance metrics are qualitatively similar.

5.1 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 for each period t against each period $s \neq t$. Aggregating across all 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 1. The same set of tests run on each mechanism (rather than aggregating across all four mechanisms) fails to find significant evidence of learning.[5] Finally, regressing the distance from the full information posterior on the period number results in a negative but insignificant coefficient on the period number, further indicating that performance does not significantly improve with experience.

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

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

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 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 1 demonstrate this result.[6]

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

5.2 The Simple Environment: Two States

The two-state environment represents a relatively simple aggregation problem where one might expect mechanisms to be more successful at approximating the full information posterior.

5.2.1 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 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 $A \succ B$ if mechanism A generates significantly higher error than B at the 10% significance level. Similarly, we define $\succeq$ by $A \succeq B$ if A has greater average error than B (regardless of the statistical significance). Since $\succ$ is not negatively transitive (it is possible to have 'not $A \succ B$' and 'not $B \succ C$' but $A \succ C$), describing the relation between mechanisms may require multiple statements. In particular, we describe the results with a set of statements such that if a statement exists of the form '$A \succeq \ldots \succ \ldots \succeq D$' then conclude that $A \succ D$. If there is no such statement, conclude that $A \not\succ D$. If '$A \succeq \ldots \succeq \ldots \succeq D$' or '$A \succeq \ldots \succ \ldots \succeq D$' exist then conclude that $A \succeq D$. For example, from the pair of statements '$A \succeq B \succeq C \succeq D$ and $A \succ C \succeq D$' we conclude that the ordering of the average errors is alphabetical and that $A \succ C$ and $A \succ D$, but not $A \succ B$.

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 2. The average error for each mechanism is reported in the second row and second column of the table; the Market Scoring Rule generates the most error on average and the Double Auction generates the least. The p-values of the pairwise Wilcoxon tests are reported in

[6] Note that a crossover design was used, where each mechanism ordering is reversed in a later session. This would allow for statistical control of any ordering effect, but no such effect has been found.

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 Market Scoring Rule is the unique mechanism that generates significantly higher error than some 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.

5.2.2 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 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 Market Scoring Rule. 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, Market Scoring Rule, 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 stagnation) 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.[7]

5.2.3 Catastrophes: Mirages

Historically, a mirage refers to a situation where a mechanism's output in a two-state environment leans towards one state when the other is the true state. In more general environments, if p^0 represents the prior distribution, p^{FI} the full information posterior given the signal s, and h the mechanism output, we define

[7] 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. Thus, these trades are not informative (or financially consequential) and are discarded from the analysis.

a mirage as any h such that $(p^{FI} - p^0) \cdot (h - p^0) < 0$. Thus, a mirage occurs when the mechanism output moves in the wrong direction from the prior, relative to the full information posterior. In this situation the mechanism has not only failed to aggregate information properly, it is misleading the market observer into believing that a different limit posterior is the maximum likelihood than she would believe had she observed the entire signal s. A graphical representation of a mirage (for both two- and eight-state environments) is provided in Figure 3.

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 Market Scoring Rule to the Pari-mutuel.[8] 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.[9]

5.2.4 Catastrophes: Confusion

Recall that, regardless of the actual signal, the full information posterior must lie in the convex hull of the limit posterior. If a mechanism's output distribution lies outside the convex hull, it cannot be rationalized as being generated from any possible signal. Consider a market observer who observes the mechanism output and generates a posterior belief about the probabilities of the events in question. Observing an output distribution outside the convex hull tells the market observer that the mechanism has somehow failed in that it failed to converge to a sensible prediction, perhaps because some individuals were irrational or did not properly employ Bayes's Law. In this case the observer can infer little if any information from the mechanism output. We refer to such a catastrophic failure as a 'confused' outcome. A graphical representation of confused outcomes (for two and eight states) is provided in Figure 4.

The third row of Table 5 displays the number of periods in which confused outcomes occur in the two-state experiments.[10] Clearly the Poll is the most frequent; using a simple binomial test we conclude that the Poll generates confused 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, Market Scoring Rule, and Pari-mutuel, respectively). Thus, our significance statement regarding confusion is 'Poll $\succ$ MSR $\succeq$ Pari $\succeq$ DblAuc'.[11]

[8] The number of mirages which are simultaneously confused outcomes (meaning the mechanism output moves outside the convex hull of the limit posterior away from the full information posterior) is 0, 1, 1, and 3 for the Double Auction, Market Scoring Rule, Pari-mutuel, and Poll, respectively.

[9] The binomial test between the Market Scoring Rule (with 14 mirages) and the Pari-mutuel generates a p-value of 0.0549.

[10] We do find that, across all mechanisms, confusion is significantly more likely to occur in the first period. No other period effects have been observed.

[11] Conditional on observing a confused outcome, the average distance between h and the convex hull is 0.024, 0.171, 0.106, and 0.052 for the Double Auction, Market Scoring Rule, Pari-mutuel, and Poll, respectively. Thus, the 'magnitude' of the confusion in the Poll is less than in the Market Scoring Rule or Pari-mutuel, though it is not clear that this measure is relevant since *all* confused outcomes lead to an inference failure, despite the magnitude.

5.2.5 Summary

In each of our four measures (error, no trade, mirages, and confusion) we found one mechanism to be uniquely bad and the others to be roughly equivalent. Specifically, the Market Scoring Rule generates the most error, the Pari-mutuel generates the most no-trade periods, the Poll is the most frequently confused, and the Market Scoring Rule 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.

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

5.3 The Complex Environment: Eight States

5.3.1 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 Market Scoring Rule and Poll generate the lowest errors on average, with no significant difference between them.

5.3.2 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.

5.3.3 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 3.

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.

5.3.4 Catastrophes: Confusion

Recall that an output distribution is labeled 'confused' 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 dimensions are fixed at the prior level. Thus, an output distribution is automatically 'confused' if either the first or last dimension differs from the prior. See Figure 4 for a simplified representation of this issue.

In practice, confusion occurs in every period under every mechanism in our eight-state experiments, so indicating confusion with a binary indicator variable is not informative. Although any confusion leads to 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 Market Scoring Rule 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 Market Scoring Rule and the Poll are less prone to large deviations from the convex hull.

An alternative way to measure the propensity for confusion 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 Market Scoring Rule and the Poll generate output distributions within ϵ of the convex hull most frequently when ϵ is small. As ϵ is increased, however, the Market Scoring Rule moves from most frequent to least frequent and the Parimutuel moves from second-least frequent to most frequent. In other words, the Market Scoring Rule 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 confusion should prefer the Pari-mutuel mechanism over the Market Scoring Rule 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.

5.3.5 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 confusion), 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 confusion.

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.

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 processes. 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 agenda; many obvious and interesting questions remain open for further 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 hypothesize that the failure of the double auction in the eight-state experiments is due primarily to the increased ratio of the number of securities to the number of traders: the ‘thin markets’ problem. Perhaps as the number of securities exceeds the number of traders, agents 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.

Another question is why the Poll, which performs relatively well in the eight-state experiments, suffers from confusion in the two-state experiments. This is particularly surprising given that the Poll's output distribution is the arithmetic mean of the distributions submitted in the final round by the three traders, so confusion in the output distribution is likely caused by either significant confusion by one trader or mild confusion by multiple traders. Examining individual-level data from our experiments may shed some light on these (and other) issues.

References

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

Robert Forsythe, Thomas R. Palfrey, and Charles R. Plott. Asset valuation in an experimental market. *Econometrica*, 50(3):537–568, May 1982.

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

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

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

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

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

Charles R. Plott, Jorgen Wit, and Winston C. Yang. Parimutuel betting markets as information aggregation devices: Experimental results. *Economic Theory*, 22:311–351, 2003.

Kevin A. Roust and Charles R. Plott. The design of information aggregation mechanisms: A two-stage parimutuel market to avoid mirages (bubbles). December 2005.

Claude E. Shannon. A mathematical theory of communication. *Bell Systems Technical Journal*, 27:379–423, 623–656, 1948.

Erik Snowberg and Justin Wolfers. Explaining the favorite-longshot bias: Is it risk-love, or misperceptions? October 2006.

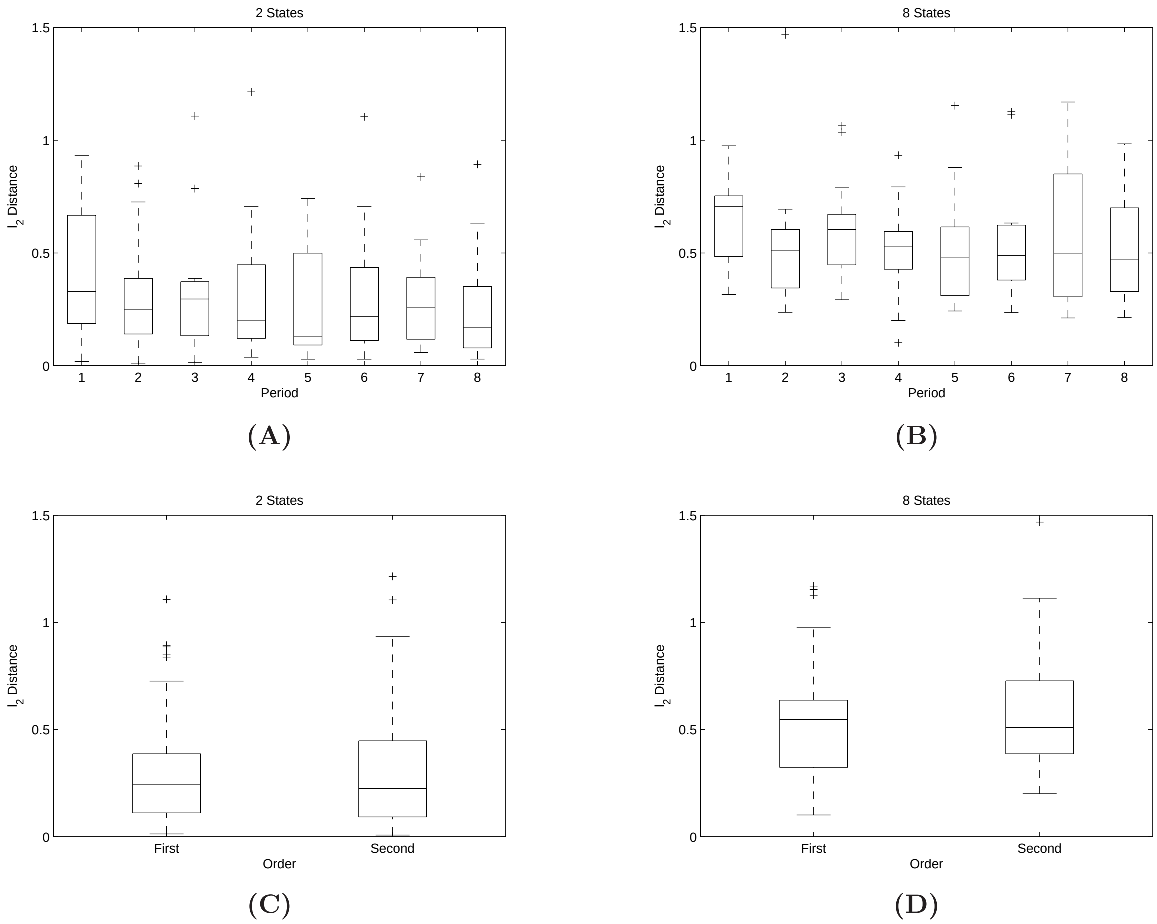

Figure 1: Box-and-whisker plots of the distance between the mechanism output distribution and the full information posterior for (A) each period in the two-state experiments, (B) each period in the eight-state experiments, (C) each mechanism ordering in the two-state experiments, and (D) each mechanism ordering in the eight-state experiments.

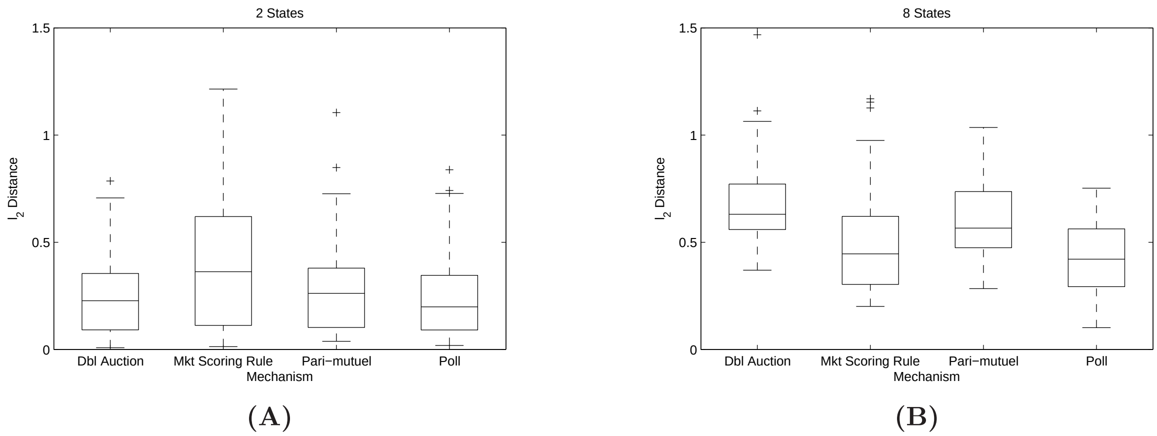

Figure 2: Box-and-whisker plots of the distance between the mechanism output distribution and the full information posterior for each mechanism in (A) the two-state experiments, and (B) the eight-state experiments.

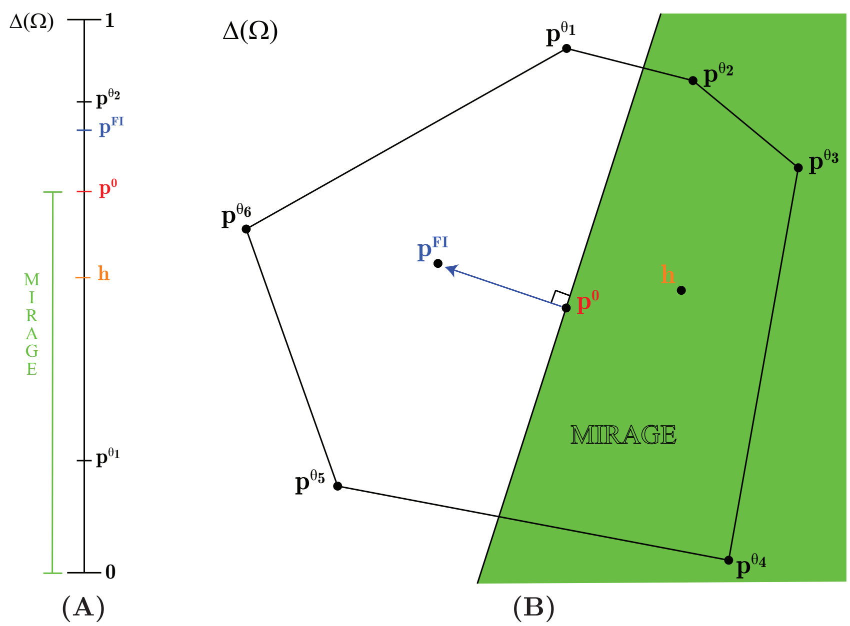

Figure 3: Mirages with (A) two states, and (B) more than two states.

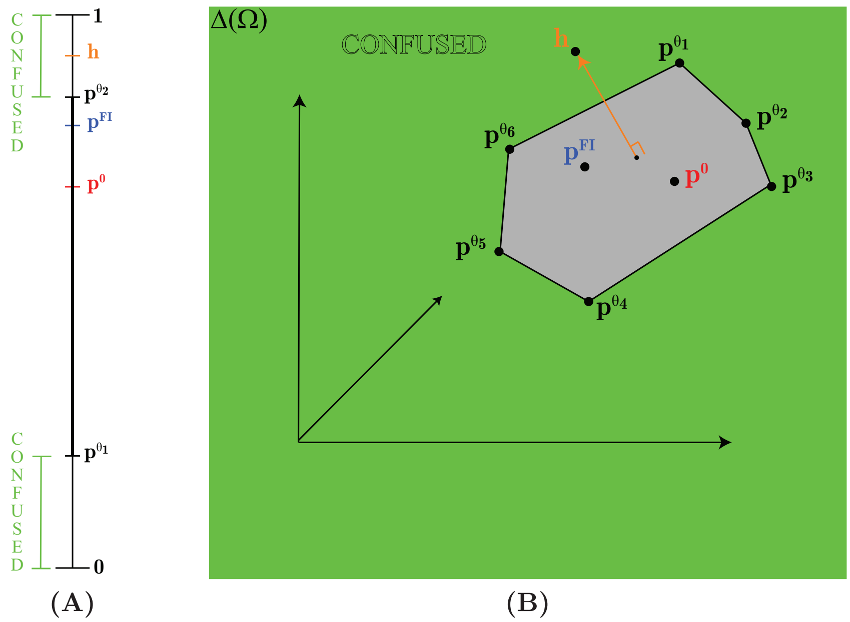

Figure 4: Confusion with (A) two states, and (B) more than two states.

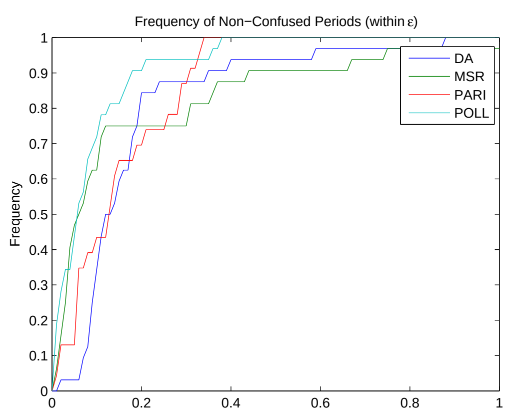

Figure 5: Frequency of periods (with trade) in which confusion is less than ϵ.

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.

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

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

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

Table 3: The experimental design.

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.

2 States	Dbl Auction	Mkt Scoring Rule	Pari-mutuel	Poll
No Trade	0	1	4	0
Confusion	5	7	6	11
Mirage	13	14	10	12
Confused 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.

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.

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 confusion 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 $(p^{FI} - p^0)$.

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.

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 confusion, 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	Confusion	Error	No Trade	Mirage	Confusion
Dbl Auction	✓	✓	✓	✓	$\times$	✓	$\times$	$\times$
Mkt Scoring Rule	$\times$	✓	$\times$	✓	✓	✓	✓	✓
Pari-mutuel	✓	$\times$	✓	✓	$\times$	$\times$	✓	$\times$
Poll	✓	✓	✓	$\times$	✓	✓	✓	✓

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