Rev. Econ. Design (2013) 17:95–128
DOI 10.1007/s10058-013-0144-z

ORIGINAL PAPER

The Gates Hillman prediction market

Abraham Othman · Tuomas Sandholm

Received: 29 December 2011 / Accepted: 16 April 2013 / Published online: 28 April 2013

Abstract The Gates Hillman prediction market (GHPM) was an internet prediction market designed to predict the opening day of the Gates and Hillman Centers, the new computer science complex at Carnegie Mellon University. Unlike a traditional continuous double auction format, the GHPM was mediated by an *automated market maker*, a central agent responsible for pricing transactions with traders over the possible opening days. The GHPM's event partition was, at the time, the largest ever elicited in any prediction market by an order of magnitude, and dealing with the market's size required new advances, including a novel span-based elicitation interface that simplified interactions with the market maker. We use the large set of identity-linked trades generated by the GHPM to examine issues of trader performance and market microstructure, including how the market both reacted to and anticipated official news releases about the building's opening day.

Keywords Prediction markets · Automated market making · Case studies · Market design

JEL Classification D4 · D7

An earlier version of this paper was presented at and appeared in the proceedings of the 11th ACM Conference on Electronic Commerce (EC 2010).

A. Othman (✉) · T. Sandholm
Computer Science Department, Carnegie Mellon University, Pittsburgh, PA, USA
e-mail: aothman@cs.cmu.edu

T. Sandholm
e-mail: sandholm@cs.cmu.edu

1 Introduction

Prediction markets are powerful tools for aggregating information (Berg et al. 2001; Wolfers and Zitzewitz 2004; Cowgill et al. 2009; Chen and Pennock 2010). A typical prediction market only generates a single point of interest; for instance, the probability that a certain candidate will win an election, or the percent of the vote that candidate will receive. Over more complex event spaces, however, these point estimates can be inappropriate. Consider a prediction market to estimate the US inflation rate over the next 5 years. Conceivably, market participants could be split between a very low estimate and a very high estimate. The resulting consensus of a middle value could be an accurate estimate of the expectation, but would be misleading to design policy around.

Recent theoretical work has indicated that eliciting interesting distribution properties (like the element that has maximum probability) is as difficult as eliciting an entire distribution (Lambert et al. 2008). This result motivates and justifies building systems that are capable of eliciting complete probability distributions over large event spaces. In this paper, we discuss the design of a market, the *Gates Hillman prediction market (GHPM)* that generated a complete distribution over a fine-grained partition of possibilities while retaining the interactivity and simplicity of a traditional market.

Fundamental to our design is an automated market maker (Hanson 2003; Chen and Pennock 2010) that offers three primary benefits. First, the market maker provides a rich form of liquidity because it guarantees that participants can make any self-selected trade at any time. Second, it allows instant pricing feedback to traders, rather than delayed, uncertain, potential feedback. A trader can always get actionable prices both on any potential trade she is considering and on the current values of the bets she currently holds. Third, the automated market maker obviates the need to match combinations of user-generated buy and sell orders—a problem that can be combinatorially complex (Fortnow et al. 2003; Chen et al. 2008)—making a large event space computationally feasible.

Also important to the success of the GHPM was the user interface of the trading platform. Given well-documented shortcomings in human reasoning, interfaces with the same expressive power in theory can perform quite differently in practice. A large event space implies that the typical probability of any single event is small, and people have great difficulty discriminating between small probabilities (e.g., Ali 1977). To solve this problem, the GHPM used a span-based interface with ternary elicitation queries, which we discuss in Sect. 2.3.

As the first test of automated market making in a large prediction market, the GHPM allowed us to discover two flaws in current automated market makers, which will help focus future design of market makers. Section 3 discusses the two flaws, *spikiness* and *liquidity insensitivity*, in detail and explores their theoretical roots.

Traditional laboratory experiments are generally small due to practical constraints like subject payments, training effort, and the viable duration of an experiment. For example, Healy et al. (2010) study behavior and prices in laboratory prediction markets in detail, but their experiment only had three traders. The Gates Hillman Prediction Market involved hundreds of traders making thousands of trades, and so provides an unusually rich data set, particularly when combined with interviews with traders about the strategies they employed. Sections 4 and 5 use the data generated by the market to examine its performance and characteristics in depth.

2 Market design

The GHPM used a raffle-ticket currency tied to real-world prizes, an automated market maker, and a novel span-based ternary elicitation interface. In the following sections we discuss each of these in turn.

2.1 Incentives and setup

Due to legal concerns, the GHPM used raffle tickets as currency rather than real money. Thanks to generous grants from Yahoo! and other sources, we secured the equivalent of about $2,500 in prizes to distribute. At the close of the market, prize selection slots were allocated randomly, in proportion to the number of tickets each user amassed. Traders then selected their prizes in descending order until all the prizes were exhausted. This gives (risk-neutral) participants the same incentives as if real money were used—unlike the approach where the best prize is given to the top trader, the second-best prize to the second-best trader, etc.

The GHPM was publicly accessible on the web at whenwillwemove.com, but trading accounts were only available to holders of Carnegie Mellon e-mail addresses. For fairness, we did not allow people with direct control over the building process (e.g., members of the building committee) to participate. A screenshot of the signup page is shown as Fig. 1.

Upon signup, each user received 20 tickets, and each week, if that user placed at least one new trade, she would receive an additional bonus of two tickets. In a market with real money, we would expect that traders more interested or knowledgeable would stake more of their personal funds in the market. However, in a fake-money setting, we do not have this option. For instance, a mechanism that asked users if they were "very interested" in the market, and promised to give them extra tickets if they answered affirmatively would obviously not be incentive compatible. The two ticket weekly bonus was intended to give more interested traders more influence in the market and to encourage traders to be more involved in the market over time. The effects of the weekly bonus on the market are discussed in detail in Sect. 5.3.

One of the most challenging parts of running a prediction market over real events is defining contracts so that it is clear which bets pay out. For example, *InTrade*, a major commercial prediction market, ran into controversy over a market it administered involving whether North Korea would test missiles by a certain date. When North Korea putatively tested missiles unsuccessfully, but the event was not officially confirmed, the market was reduced to a squabble over definitions. We set out to study when the Computer Science Department would move to its new home in the Gates and Hillman Centers (GHC), but *move* is a vague term. Does it indicate boxes being moved? Some people occupying new offices? The last person occupying a new office? The parking garage being open? From discussions with Prof. Guy Blelloch, the head of the building committee, we settled on using "the earliest date on which at least 50 % of the occupiable space of the GHC receives a temporary occupancy permit". Tempo-

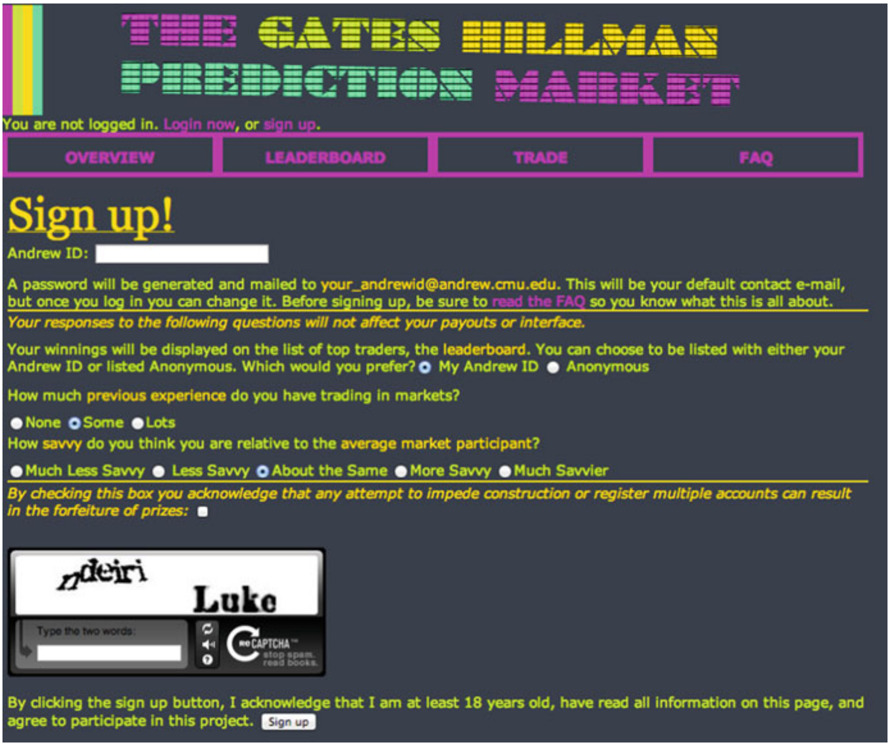

Fig. 1 The signup page for the GHPM. *Andrew IDs* are unique usernames tied to e-mail addresses that are granted to Carnegie Mellon affiliates (e.g., the Andrew ID "aothman" corresponds to an e-mail address of aothman@andrew.cmu.edu). Trader performance relative to their self-reported levels of sophistication is discussed in Sect. 5.2

rary occupancy permits are publicly issued and verifiable, must be granted before the building is occupied, and are normally issued immediately preceding occupancy (as was the case in the GHC).

The market was active from September 4, 2008 to August 7, 2009. On this latter date, the GHC received its first occupancy permit, which covered slightly over 50 % of the space in the building. The price of a contract for August 7, 2009 converged to 1 about 5 h before the public announcement that the building had received its permit.

In total, 210 people registered to trade and 169 people placed at least one trade. A total of 39,842 bets were placed with the market maker, with about two-thirds of the trades in the market being placed by a single trading bot (further discussed in Sect. 5.5). Following the conclusion of the market, we conducted recorded interviews with traders we deemed interesting about their strategies and their impressions of the GHPM. Excerpts of some of these conversations appear in Sect. 4.

2.2 Automated market maker

In this section, we provide a brief discussion of the automated market maker concept originated by Hanson (2003), and how we applied it to the specific setting of the GHPM. This idea behind automated market making has been widely applied in practice, including at *Inkling Markets*, a prediction market startup company, and (before its recent demise) *Tradesports*, a sports betting prediction market company.

We began by partitioning the event space into $n = 365$ events, one for each day from April 2, 2009 to March 30, 2010 with the addition of "April 1, 2009 and everything before" and "March 31, 2010 and everything after", to completely cover the space of opening days. At the time, the GHPM was by far the largest market (by event partition size) ever conducted. The largest prior prediction markets fielded in practice were, to our knowledge, markets over candidates for political nominations, where as many as 20 candidates could have contracts (of course, only a handful of candidates in these markets are actively traded). Previous laboratory studies have involved limited trials with as many as 256 events (Ledyard et al. 2009), but those studies involved very few traders and are not commensurable with the GHPM, which was designed as a publicly visible mechanism. Since the GHPM concluded, markets with much larger event spaces have been fielded. *Predictalot*, a product of Yahoo! Research, was a public prediction market that fielded bets on the 2010 and 2011 NCAA men's basketball tournament, a setting with 2^{63} events. As the first large-scale test of automated market making, the GHPM is the bridge between smaller, human-mediated markets and the later development of exponentially larger combinatorial markets, mediated with automated market makers.

An automated market maker, or at least some form of automated pricing, becomes a necessity as the number of events grows large. Otherwise, traders with differing views would simply have no way of matching their orders. With only two events, a bet for one event serves to match against a bet for the other. Consequently, traders submitting bets have no problem finding counterparties: they are just traders betting on the other event. But with hundreds of events, a bet for one event serves to match against the *set* of every other event. Two traders with divergent bets, say, one who believes the building will open early and the other who believes the building will open late, will not directly be each others' counterparties. Instead, the market will only clear if a whole set of such orders—possibly consisting of hundreds of distinct orders—that span over the totality of events can be matched. So, even clearing a single order could require hundreds of competitive orders placed and waiting. In fact, the problem is even worse than what has been described here. In practice, if traders do not see feedback on their bets, they will likely withdraw from the market entirely, meaning that the hundreds of primed orders required to clear the market will never be present, and the market will fail. Furthermore, if agents are allowed to submit orders on arbitrary combinations of events, the market clearing problem becomes NP-hard (Fortnow et al. 2003). Automated market makers solve these problems by automating a counterparty to step in and price bets for traders. The tradeoff is that this automated agent can, and generally will, run at a loss. This loss can be though of as a subsidy to elicit their information (Pennock and Sami 2007).

For the GHPM we applied the most widely-used automated market maker for prediction markets, the *logarithmic market scoring rule (LMSR)*, originally designed by Hanson (2003, 2007).[1] The market maker operates according to a cost function $C : \Re^n \mapsto \Re$, which maps a vector of quantities q to a scalar representing how much money has been paid into the system. Each entry q_i in the vector q represents how much money is to be paid out if the ith event is realized. For instance, consider a market over two events, corresponding to whether the Red Sox or Yankees will win their next baseball game. If the market maker has taken bets that sum to paying out five dollars if the Red Sox win and three dollars if the Yankees win, then $q = (5, 3)$. If a trader wants to make an additional bet that pays out one dollar if the Red Sox win, then the market maker would charge $C((6, 3)) - C((5, 3))$.

The cost function for the LMSR is

$$C(q) = b \log\left(\sum_i \exp(q_i/b)\right)$$

where $b > 0$ is a constant fixed *a priori* by the market administrator. As Pennock and Sami (2007) discuss, the b parameter can be thought of as a measure of market liquidity, where higher values represent markets less affected by small bets. In the GHPM we fixed $b = 32$, and since the LMSR has worst-case loss of $b \log n$ (Hanson 2003, 2007), at most about 190 surplus tickets would be won from the market maker by participating traders. (This is indeed the amount actually transferred from the market maker to the participants because probability mass converged to the correct day before the market ended.) The *ad hoc* nature of selecting the liquidity parameter is an intrinsic feature of using the LMSR, as it comes with little guidance for market administrators (Othman et al. 2010). The parameter we selected turned out to be too small, which led to some problems in the market (see Sects. 3.2 and 5.3).

Prices are defined by the gradient of the cost function, so that

$$p_i(q) = \frac{\exp(q_i/b)}{\sum_j \exp(q_j/b)}$$

is the price of the ith event. We call these p_i *pricing rules*. The prices can also be directly thought of as event probabilities, because they define a probability distribution over the event space: they sum to unity, are non-negative, and exist for any set of events.

To better understand how the market maker works, consider the following simulated interaction. Imagine, instead of fielding bets on when the new building would open, that we were fielding bets on whether the Red Sox or Yankees will win their next baseball game. For clarity, we will have the market maker be the same as the one we used in our market: The LMSR with liquidity parameter $b = 32$. Again, suppose that the market maker has taken bets that sum to paying out five dollars if the Red Sox win

[1] Recent years have seen the introduction of many new market makers that operate similarly to the LMSR in terms of qualitative properties but that have different price responses. These include the convex risk measures of Ben-Tal and Teboulle (2007) and Agrawal et al. (2009) and the constant-utility cost functions of Chen and Pennock (2007) and more sophisticated agents that attempt to mitigate the market maker's subsidy (Othman and Sandholm 2012). Automated market makers are also used in more theoretical work involving market microstructure and simulation (Ostrovsky 2009; Othman and Sandholm 2010; Chakraborty et al. 2011).

and three dollars if the Yankees win, so that $q = (5, 3)$. Now, a trader approaches the market maker with the intention of placing a bet that would pay out one dollar if the Red Sox win (and zero dollars if the Yankees win). This bet would change the market maker's quantity vector to (6, 3). The market maker computes the following to price the bet:

$$C((6, 3)) - C((5, 3)) \approx 0.5195$$

and so would quote the trader a price of about 52 cents. Before the trader placed the bet, the price of the Red Sox winning is

$$\frac{\exp(5/32)}{\exp(5/32) + \exp(3/32)} \approx 0.5156$$

and if the trader accepts the bet, the price of the Red Sox winning would be

$$\frac{\exp(6/32)}{\exp(6/32) + \exp(3/32)} \approx 0.5234$$

Observe that placing the bet on the Red Sox increases the market maker's price on the Red Sox winning. This example provides motivation for the well-known result that myopic, risk-neutral traders without capital constraints will interact with the LMSR until the market maker's prices reflect the trader's beliefs (Hanson 2003, 2007).

Since the market ran to August 7, 2009 but the event space started with April 1, 2009, for the last several months of the market some of the events corresponded to days in the past that had expired worthless. For instance, the event "The buildings open on June 10, 2009" was still incorporated into the cost function on June 11, 2009—even though that earlier day had come and gone without the buildings opening. Rather than incorporate the notion that these events were expired into the market maker, we did not change the market maker's pricing at all. We observed that as dates expired, speculators would drive their prices down to zero (either by directly betting against those days or by wagering tickets on days that were still possible). Our daily data snapshots indicate that the total amount of probability weight on expired events never exceeded 1 % and was typically negligibly small (<0.05 %, which would have been displayed as 0 % to traders).

2.3 Span-based elicitation with ternary queries

In this section, we present the novel elicitation mechanism used in the GHPM. A similar interface was developed independently and contemporaneously by Yahoo! Research for Yoopick (Goel et al. 2008), an application for wagering on point spreads in sporting events that ran on the social network Facebook.

The major problem in implementing fine-grained markets in practice is one of elicitation: they are too fine for people to make reliable point-wise estimates. Consider the GHPM, which is divided into 365 separate contracts, each representing a day of a year. Under a traditional interaction model, traders would act over individual contracts, specifying their actionable beliefs over each day. But with 365 separate contracts, the average estimate of each event is less than 0.3 %. People have great difficulty reliably distinguishing between such small probabilities (Ali 1977), and problems estimating low-probability events have been observed in prediction markets (Wolfers and Zitzewitz 2006).

We solve this problem by simple span-based elicitation, which makes estimation of probabilities easy for users. In our system, the user can select a related set of events and gauge the probability for the entire set. Spans are a natural way of thinking about large sets of discrete events: people group months into years, minutes into hours, and group numbers by thousands, millions, or billions. The key here is that spans use the concept of distance between events that is intrinsic to the setting.

For example, let the market be at state $q^{\mathbf{0}} = \{q_1^0, \ldots, q_n^0\}$. (Recall that these are the quantities that will be paid out if each of the respective states is realized.) A user's interaction begins with the selection of an interval from indices s to t. This partitions the indices into (at most) three segments of the contract space: $[1, s)$, $[s, t]$, and $(t, n]$. The user then specifies an amount, r, to risk. Our market maker proceeds to offer the following alternatives to the user:

- The "for" bet. The agent bets *for* the event to occur within the contracts $[s, t]$. The user's payoff if he is correct, π_f, satisfies

$$C\left(q_1^0, \ldots, q_{s-1}^0, q_s^0 + \pi_f, \ldots, q_t^0 + \pi_f, q_{t+1}^0, \ldots, q_n^0\right) = C(q^{\mathbf{0}}) + r$$

- The "against" bet. The agent bets *against* the event occurring within the contracts $[s, t]$. The user's payoff if he is correct, π_b, satisfies

$$C\left(q_1^0 + \pi_b, \ldots, q_{s-1}^0 + \pi_b, q_s^0, \ldots, q_t^0, q_{t+1}^0 + \pi_b, \ldots, q_n^0 + \pi_b\right) = C(q^{\mathbf{0}}) + r$$

As long as the cost function is strictly increasing, as is the case for the LMSR, these values uniquely exist. However, solving for π_f and π_b is not generally possible in closed form. These equations can be solved numerically using, for example, Newton's method. Depending on the specific cost function and numerical solution method, there may be issues with solution instability that should be addressed; for instance, the GHPM used Newton's method with a bounded step size at each iteration to discourage divergence.

Given a selected set of events, the simplest way to represent a bet for that set is to have each event in the set pay out an identical amount if the event is realized, as we do in the two equations above. This simplicity means we can significantly condense the language we use when eliciting a wager from an agent. Instead of asking a user whether he would accept an n-dimensional payout vector, we need only present a single value to the user. A screenshot of the elicitation process in the GHPM appears as Fig. 2. (To get this display, the user would have selected a span from a displayed probability distribution over opening days, as in Fig. 3, and their number of tickets risked from a slider. The calculated displays are re-computed when the user changes either the number of tickets wagered or the span.)

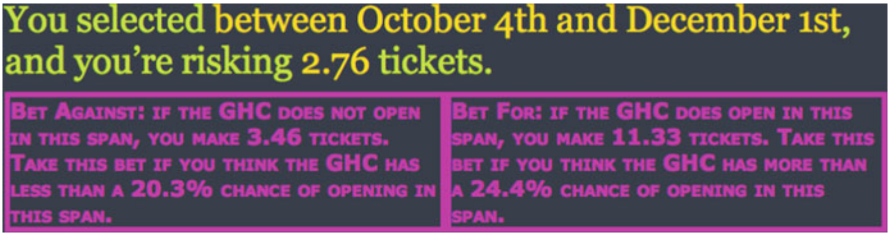

Fig. 2 A screenshot of the elicitation query for a user-selected span in the GHPM. The query is ternary because it partitions the user's probability assessment into three parts. *The GHC* is the Gates and Hillman Centers, the new computer science buildings at Carnegie Mellon. Because of legal concerns, the market used raffle tickets rather than money

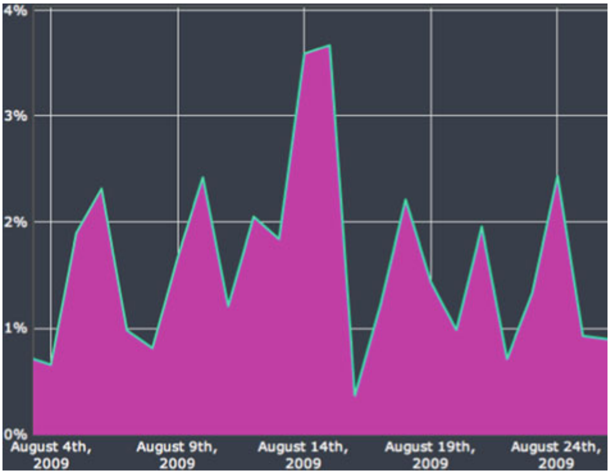

Fig. 3 A screenshot of the GHPM that shows the spikiness of prices. The x-axis ranges over a set of potential opening days. The y-axis displays prices (as percentages; e.g., 1 % means a price of 0.01)

Yahoo! Research's Yoopick does not have "against" bets, but the GHPM does. From discussions with traders in the GHPM, against bets were used frequently to bet against specific (single) contracts they feel are overvalued. Several successful traders had a portfolio consisting solely of bets against a large number of single contracts. The success of these traders was likely a combination of the misjudging of small probabilities by other traders as well as the spiky price phenomenon discussed in the next section.

There are several relevant pieces of information the market administrator could provide the users for each potential bet:

- The agent's direct payout if he is correct, π_f (or π_b). This is the amount a trader wins if he bets on a span including the event that occurs and he holds the contract through to expiry. Both Yoopick and the GHPM display this information.
- The *averaged* payout probability on the span, r/π_f or $1 - r/\pi_b$. This value is the actual odds at which a bet is being made. The GHPM displays this as a *ternary* (three-way) query, where agents can select whether their probability estimate lies in one of three partitions, as in Fig. 2. (Yoopick does not display this information.)

Faced with the ternary query, the user selects whether his probability for the span is less than $1 - r/\pi_b$, greater than r/π_f, or in-between. (If the prices are increasing in quantities, as they are in the LMSR, then $r/\pi_f \geq 1 - r/\pi_b$, with equality in the limit as $r \to 0$.) If an agent's belief lies in the middle partition, presumably they could reduce their bet size or find another span on which to wager.
- The *marginal* payout probability, which is the sum of the prices on the relevant span after the π_f or π_b of additional quantity. Since agents who are acting straightforwardly will not want to move marginal prices beyond their private valuation, marginal prices could be more informative to decision making. Neither the GHPM nor Yoopick displays this information. Early trials of the GHPM included marginal prices in the interaction interface, but testers found the information confusing when combined with the averaged payout probabilities and so we removed the marginal payout probabilities from later versions of the interface. Even though they were not explicit in the interface, sophisticated traders could still produce marginal prices either by explicitly knowing the pricing rule or by making small tweaks in the number of tickets risked and observing how prices changed. We feel that for a market exclusively populated by mathematically adept traders, explicit marginal prices would be a helpful tool.

Finally, even though it simplifies interactions, the span-based elicitation scheme is arbitrarily expressive. If the users are sophisticated enough to make discriminating judgments over small probabilities, to the point that they can express their actionable beliefs over every contract, then they can still express this sophistication using spans—e.g., by trading spans that contain only one element (one day in the case of the GHPM).

3 Problems revealed

There are two key findings from our study. The first is the large data set of trades, which we analyze in Sects. 4 and 5. The second is the discovery of two real-world flaws in the automated market-making concept. These were the *spikiness* inherent in prices and the *liquidity insensitivity* that made prices in the later stages of the market change too much. We proceed to discuss these flaws in the next two sections, respectively.

3.1 Spikiness of prices across similar events

A phenomenon that quickly arose in the GHPM was how spiky the prices were across events at any snapshot in time. There was extraordinary local volatility between days that one expects should have approximately the same probability. This volatility is far more than could be expected from a rational standpoint—e.g., weekends and holidays could be expected to have much lower probability than weekdays—and it persisted even in the presence of profit-driven traders whose inefficiency-exploiting actions made for less pronounced, but still evident, spikes. (These traders' strategies are discussed in detail in Sect. 4.2.) Figure 3 is a screenshot of the live GHPM where spiky prices are evident. Spikiness here refers to the exaggerated sawtooth pattern of prices in a fixed snapshot in time, not to how prices moved or changed over time. It is a distinct phenomenon from both real markets with sharply rising or falling prices (e.g., electricity spot markets) or from the "mirages" observed in laboratory studies of the LMSR (Healy et al. 2010).

3.1.1 Spikiness in theory

Why did spikiness occur? Is there an automated market maker with a different pricing rule that would have resulted in a market where prices were not spiky? In this section, we show that a large class of market makers will tend to induce spiky prices. Specifically, our result concerns pricing rules which satisfy three conditions: differentiability, non-negativity, and a technical condition we call *pairwise unboundedness*.

Definition 1 When $n \geq 3$, a pricing rule is *pairwise unbounded* if, when enough is bet against any pair of events, the prices on both of those events goes to zero. Formally, let the set $\mathbf{I}$ consist of any two events i and j, and let $\mathbf{I}$ be the indicator vector for the set $\mathbf{I}$, so that $\mathbf{I}_i = \mathbf{I}_j = 1$ and $\mathbf{I}_{\neq i \vee j} = 0$ otherwise. Then for any x

$$\lim_{k\to\infty} p_i(x - k\mathbf{I}) = \lim_{k\to\infty} p_j(x - k\mathbf{I}) = 0$$

Of the three properties, non-negativity and pairwise unboundedness are the most natural. A negative price would make the marginal prices not form a probability distribution, and would imply a trader could arbitrage the market maker by "buying" the event with a negative price. The pairwise unbounded condition is also natural, because a market maker that is not pairwise unbounded will keep non-vanishing prices on pairs of events no matter how much is bet against them by traders. To our knowledge, every market maker in the literature satisfies both non-negativity and pairwise unboundedness. In contrast, there do exist market makers that do not have a differentiable price response. However, a smooth price response is desirable for interacting traders (Othman and Sandholm 2011), and the bulk of market makers from the literature do have differentiable prices. It is easy to see that the LMSR satisfies all three conditions.

What does it mean for a market maker to produce spiky prices? From examining transactions in the GHPM, we found that spiky prices arose most often from tiny differences in the prices of nearby days being amplified by a bet span that included both of them. These tiny initial price differences seem to arise endogenously from traders selecting slightly different intervals to bet on, although there is no guarantee that they will be present. With this in mind, consider a bet placed for the interval $\mathbf{I} = \{i \text{ and } j\}$, when $p_i > p_j$. Recall that the derivative of a differentiable function $g : \Re^n \mapsto \Re$ along a vector x (the *vector derivative*) is given by the scalar $\nabla_x = \nabla g \cdot x$. The vector derivative $\nabla_{\mathbf{I}}$ of the price functions p_i and p_j is

$$\nabla_{\mathbf{I}} p_i = \nabla_i p_i + \nabla_j p_i$$

and

$$\nabla_{\mathbf{I}} p_j = \nabla_j p_j + \nabla_i p_j$$

The market maker has spike-inducing behavior if betting on the bundle that consists of both i and j amplifies the difference between the prices on those events, that is, if

$$\nabla_{\mathbf{I}} p_i > \nabla_{\mathbf{I}} p_j$$

By the symmetry of second derivatives, $\nabla_j p_i = \nabla_i p_j$. Therefore we have spike-inducing behavior for this bet if

$$\nabla_i p_i > \nabla_j p_j$$

This leads us to the following definition.

Definition 2 In a setting where $n \geq 3$, a market maker *induces spikes* if

$$\nabla_i p_i(x) > \nabla_j p_j(x)$$

for some x and some i and j such that

$$p_i(x) > p_j(x)$$

Theorem 1 *Every differentiable, non-negative, and pairwise unbounded pricing rule induces spikes.*

Proof Suppose there is a pricing rule satisfying the three conditions that does not induce spikes. Then, for every x, i, j tuple we have

$$\nabla_i p_i(x) \leq \nabla_j p_j(x)$$

Select some x, some indices i and j, and some $\epsilon > 0$ such that

$$p_i(x) \geq p_j(x) + \epsilon$$

Because the market maker has differentiable, pairwise unbounded prices, such an x, i, j tuple must exist.

Now consider the vector $\mathbf{I}$ which will be the indicator vector for indices i and j. Because the market maker does not induce spikes we have that, for $k \geq 0$

$$p_i(x - k\mathbf{I}) \geq p_j(x - k\mathbf{I}) + \epsilon$$

But because the market maker has pairwise unbounded prices,

$$\lim_{k \to \infty} p_i(x - k\mathbf{I}) = 0$$

Consequently there exists some $K > 0$ such that

$$p_i(x - K\mathbf{I}) \leq \epsilon/2$$

But then at this K,

$$p_j(x - K\mathbf{I}) \leq -\epsilon/2$$

so the market maker has prices that are negative, but this is a contradiction, because we assumed the pricing rule was non-negative. □

3.1.2 The impact of spikiness

Traders were aware of spikiness and this knowledge affected their behavior. In Sect. 4, we discuss and analyze interviews with traders which suggest that spikiness played a large role in determining the way that some traders behaved in the GHPM.

Spiky prices are a problem because they create a disconnect between the user and the elicitation process. Users feel that the spiky prices they observe after interacting with the market maker do not reflect their actual beliefs. This is because users agree only to a specified bet rather than to an explicit specification of prices after their interaction. Moreover, because the difference between spiky prices and (putatively) efficient prices is so small, traders have little incentive to tie up their capital in making small bets to correct spikiness; there is almost certainly another interval where their actionable beliefs diverge more from posted prices. Our interview with Brian, a PhD student in the Machine Learning Department and the market's best-performing trader, was informative. He described a sophisticated strategy where he would check the future prospects of his current holdings against what he viewed as a risk-free rate of return—for instance, by betting against the building opening on a weekend. If the risk-free rate of return was higher, he would sell his in-the-money holdings and buy into the risk-free asset. So once a spike is small enough, damping it out can be less lucrative than other opportunities. (This argument also implies that spikiness could be diminished by supplying traders with more capital.)

Finally, to bet against a spike, a trader accepts an equal payout on every other day. But increasing the quantity on events by the same amount is what caused spiky prices in the first place. Put another way, betting against a spiky price will have the tendency to create spiky prices elsewhere in the probability distribution. Consequently, when using a differentiable market maker, while it might be possible for savvy traders to diminish spikiness, it seems unlikely that it could ever be fully eliminated.

3.2 Liquidity insensitivity

Recall that the cost function used in the GHPM was

$$C(q) = b \log\left(\sum_i \exp(q_i/b)\right)$$

and that prices are given by the gradient of this function

$$p_i(q) = \frac{\exp q_i/b}{\sum_j \exp q_j/b}$$

Defining $\mathbf{1} \equiv (1, 1, \ldots, 1)$, it is evident by inspection that

$$p_i(q) = p_i(q + \alpha\mathbf{1})$$

for scalar α. Hanson (2003) and Chen and Pennock (2007) present this relation as a property of any arbitrage-free market maker, because it ensures that

$$C(q + \alpha\mathbf{1}) = C(q) + \alpha$$

so that buying a guaranteed return of α regardless of the realized outcome should cost α.

A practical interpretation of this result is that the market maker is *liquidity insensitive*, so that quoted responses do not respond to the level of activity seen in the market. This implies that prices change exactly the same amount for a one dollar bet placed at the start of the market (say, at $q = (0, 0, \ldots, 0)$) as after the market maker has matched millions of dollars ($q = (1{,}000{,}000, \ldots, 1{,}000{,}000)$).

This is not the way we think of markets in the real world, operated by humans, as working. As markets grow larger with more frequent trading, they become deeper so that small bets have vanishingly small impact on prices. When using the LMSR, because the liquidity parameter b is a constant, the market's reaction to bets at increasing levels of volume is constant, too.

3.2.1 Impact on trader behavior

Liquidity insensitivity had an impact on traders in the GHPM, but unlike spikiness, which was publicly visible and a source of frequent consternation in our interviews, it appears that only the most active and savvy traders were aware of liquidity insensitivity. Brian, the market's best trader, said this about the way he approached the market in its final weeks:

> One thing I noticed was that at the end, these small bets would still make big jumps in the prices. So I would try to keep the amount that I bet really small . . . to try and minimize what would happen to the prices.

So, at least the savviest traders were aware of the disconnect between the automated market maker and the way a traditional market would function.

3.2.2 Relation to spikiness

Although spikiness and liquidity insensitivity appear quite different, they are actually related. A market maker that is sensitive to liquidity would be able to temper spikiness, because in more liquid (deeper) markets, the market maker could move prices less per each dollar invested. Since spikes are the product of discrepancies in the amount that prices move, if prices move less, spikiness will be diminished.

4 Effective trader strategies

In this section, we discuss the strategies employed by profitable traders. We found these strategies can be grouped into three categories, *spike dampening, relative smoothing*,

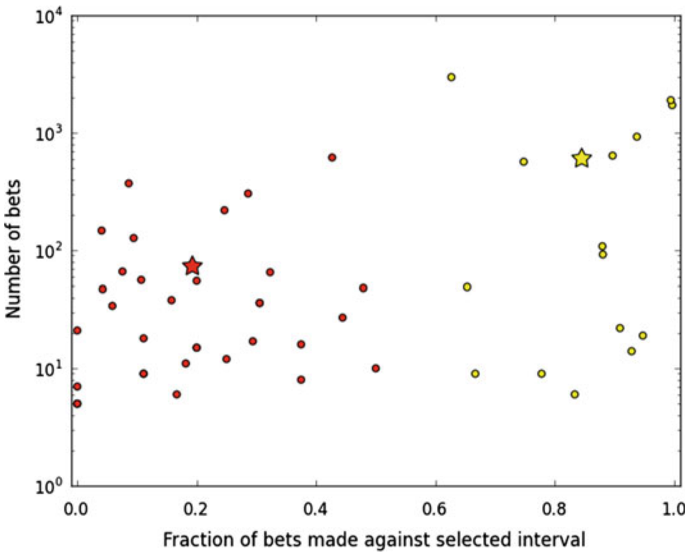

Fig. 4 A cluster analysis of profitable traders. The y-axis is log-scaled

and *information gathering*. We begin by performing a cluster analysis of the set of profitable traders to identify the qualitative attributes of successful strategies.

4.1 Cluster analysis

There were 49 traders who ended the market with more than 20 tickets and who made at least five trades (the median number over the whole set of traders). Consider grouping those traders along two criteria: First, the fraction of bets they made which were negative (i.e., *against* the span they selected, as in Fig. 2), and the number of total trades they made. Figure 4 shows the results of clustering these successful traders into two groups.

The large stars show the centroid of each group, suggesting two groups of traders: one that made a large number of negative bets, and another that made roughly an order of magnitude fewer bets, most of them positive.

The negative and positive bets themselves were also quite different. The mean negative bet was over 16.3 days and the median negative bet was on just 1 day. The mean positive bet was over 20.3 days and the median positive bet was on 4 days. (Here, the means are much larger than the medians because the samples were skewed by large outliers.) So the negative bets were over much smaller intervals than positive bets.

This analysis suggests that successful traders followed one of two distinct strategies. One group of traders made large numbers of negative bets over very small intervals; and the other made smaller numbers of positive bets on larger intervals. We call the former strategy *spike dampening* and the latter strategy *relative smoothing*. We proceed to discuss these strategies in the context of interviews with key traders we conducted immediately after the market concluded.

4.2 Spike dampening

Recall that in Sect. 3.1.1 we discussed how and why the observed market prices were spiky, with unjustifiably large variance in the prices of putatively similar individual days. Several successful traders based their strategies entirely around betting against spikes. Rob, a PhD candidate in the Computer Science Department, ended with about 256 tickets and finished in fourth place overall. In our interview with him, he described his strategy as follows:

> I knew that the market was presumably figuring out the probabilities of events, and early on, those predictions were *very* uneven. I supposed some people were setting all their money down on a single day or small set of days, and that this was causing the probability graph to be very "spiky". I bet against the spikes.
>
> Presuming (and I was correct) that as new people entered the market, the spikes would change radically and I'd cash out on the old spikes (making money) and bet against the new spikes.

Of course, on the other side of Rob's actions were traders like Jeff (a pseudonym). Jeff is another PhD student in the Computer Science Department with a background in finance; he worked as a quantitative analyst at a hedge fund before coming to graduate school. A frequent trader, Jeff finished with enough tickets to place himself in the top 15 traders. Of his experience, he said:

> It seemed like every time I would make a trade the value [of the bet] would fall a little bit . . . it was frustrating, like everything I was doing was wrong.

Jeff's bets would fall in value because they would create spikes, which speculators like Rob would quickly sell.

4.3 Relative smoothing

While spike dampening involves betting *against* spikes, relative smoothing involved betting *for* undervalued spans. Alex, a Mathematics PhD student who was one of the market's top traders, explained his strategy as follows:

> My main strategy was to purchase time periods that were given $<10\,\%$ opening probability by the market yet I felt had at least a 25 or 30 % chance of containing the opening date.
>
> Most of my gains came as the market started to believe in the summer opening. After moving up to the top five and then into the top two on the leaderboard, I was actually pretty conservative. I sold many of my successful positions for hefty gains and sat on a lot of tickets. At this point, I didn't feel like I had a lot more information that the market so I made small opportunistic trades but held on to most of the gains.

Brian, the market's top trader, employed both spike dampening as well as relative smoothing. He said:

> I was mostly just betting against [spikes], but I'd also bet for anything that was abnormally low probability . . . I was kind of like a regularizer.

One example Brian gave of an opportunity like this was a series of 3 weeks, the first and third of which have higher probability than the middle one. Brian would then buy the middle span, wait for the prices to converge, and then sell at a profit.

The strategy employed by the trading bot that made the majority of the market's trades (further discussed in Sect. 5.5) was also designed to smooth market prices. That bot attempted to fit a mixture of Gaussian distributions to the current prices, and trade off of deviations from the fit.

4.4 Information gathering

One strategy that was not identified in the cluster analysis but emerged from interviews with traders was *information gathering*. Consider that both spike dampening and relative smoothing are contrarian strategies—betting against prevailing price trends and waiting for further price movements to validate their trades. Because these strategies are both based around the relative prices of various spans (without regard to what the days actually represented), they can be considered *technical* trading strategies. In contrast, information gathering attempts to find the actual values associated with spans.

Elie, a PhD student in the Computer Science Department, invested a great deal of effort in finding out the real opening day. He became a regular at the construction site, getting updates on progress through discussions with the lead foreman, and even got the cell phone number of the building inspector who issued the temporary occupancy permit. On the morning of August 7th, the correct opening day, he visited the work site and got confirmation that the inspector had indeed signed off on the occupancy permit. He pushed the price of August 7th very high, to about 50 % of the probability mass, and tried to get spike-dampening traders to move against him so he could extract additional profit. When, instead, this high price held, he drove the price up to almost 100 % of the probability mass. Elie was the trader that ended up with the highest Information Addition Ratio (see Sect. 5.6.3), a measure of how much traders raised the price of the correct opening day. One of the desirable properties of markets is that they reward traders that acquire good information, and indeed information acquisition was extremely profitable for Elie. In the last several weeks of trading, he moved from being in around 100th place to finish as the fifth-ranked trader in the market.

4.5 Frequency of technical trading

It is useful to quantity how frequently subjects employed strategies of a technical nature. To that end, in this section we investigate the frequency of trades made against single days (presumably spikes) in the market. We selected these trades because, while they tend to improve the efficiency of the market, they are also more obviously of the technical rather than the informational type. A purely technical relative smoothing trade is much harder to distinguish from an informational trade; is a trader betting on

Table 1 Counts of traders by inclusion in the successful group and employment of the "bet against a single day" technical strategy

Made a bet against a single day?	Successful traders	Other traders
Yes	14	5
No	35	115

a "cheap" three-day span because it is inexpensive, or because the trader genuinely thinks the building will open in that span? By contrast, spike dampening trades have much more limited upside and lend themselves to rote, mechanical application (e.g., "If I see a spike, sell against it") rather than informed deliberation.

We begin by partitioning the traders into two sets; our 49 successful traders (who finished with more than 20 tickets and made five or more trades) and the 120 other traders (who either made fewer than five trades or ended the market with fewer than 20 tickets, or both). 14 of the 49 successful traders (29 %) made at least one bet against a single day, but just 5 of the 120 other traders (4 %) made at least one bet against a single day. These results are shown in Table 1.

The probability of a trader being in the successful group conditional on them making a bet against a single day is 74 %, despite successful traders making up only 29 % of the trading population. This result suggests that, although the majority of successful traders did not employ a technical spike-eliminating trading strategy, traders employing this technical strategy were more likely to be successful than not.

5 Analysis of empirical market performance

The large data set of trades linked to user accounts is a valuable product of the GHPM. In this section, we use this data set to explore questions related to trader behavior and performance, and its impact on prices and information aggregation.

5.1 The GHPM reacted to official communications, but correctly anticipated a delayed opening

The market had a complicated reaction to official communications about the opening day. We provide evidence that the market reacted to official communications, but that it correctly anticipated an officially unexpected delay in opening. (Recall that the GHPM was designed to test existing automated market makers over large event spaces, large numbers of participants, and long market durations. It was *not* designed to out-predict technical project completion forecasts from experts.)

Figure 5 shows how the distribution of prices changed over time and Table 2 shows the officially communicated moving dates. As we explained earlier, the moving day provides an upper bound on the issuance of the occupancy permit because people are not allowed to move into a building without a permit.

We can provide a rough narrative of the market from these two sources. Following some initial skepticism, market prices moved towards the correct prediction, becoming very prescient by the end of November. By then, the exterior framing of the buildings was complete. Over the next several months, the outside appearance of the buildings

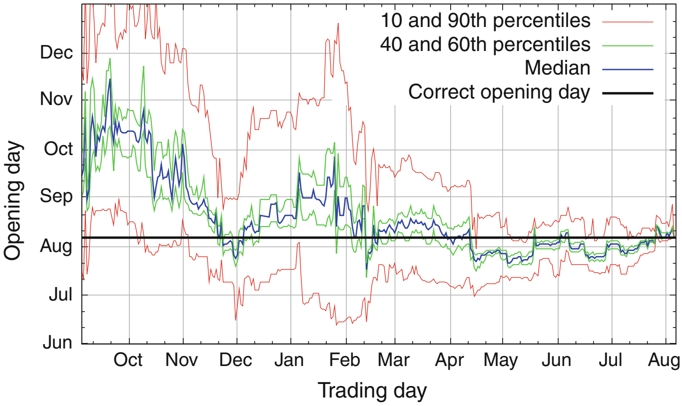

Fig. 5 The set of prices offered by the market maker corresponds to a probability distribution. This figure shows the percentile *curves* of the probability distribution. The x-axis ranges over the trading days of the market, while the y-axis ranges over the possible opening days (the contracts in the market). The market actually spanned opening dates from April 1, 2009 to March 31, 2010; the y-axis is truncated here for clarity

Table 2 Officially communicated moving dates

Date of communication	Moving day	Medium
October 15, 2008	July	Blog
February 14, 2009	August 3rd	E-mail
July 23, 2009	August 3rd	E-mail
July 28, 2009	Approximately August 10th	E-mail

did not improve measurably, and there were no official communications during this period. Prices reflected this seeming lack of progress.

The weather may have further reinforced traders' beliefs in a delay; the winter of 2008–2009 was particularly cold in Pittsburgh and featured the lowest temperatures in 15 years. Pittsburgh's average temperature in January 2009 was 22°, 6° colder than the historic average of 28°. As Fig. 6 shows, the market's probabilities for the building opening in early August peaked in late November and steadily fell throughout December and January.

Did market prices anticipate or lag public disclosures? To test this, we simulated the performance of a trader with inside information of the public announcements. (Recall that the members of the building committee that made these announcements were not allowed to trade in the GHPM.) How well would the official-information trader have done?

We considered two different schemes for how such an inside trader could operate:

- In the *Sell-quickly* scheme, the official trader spends some fraction of his wealth the day before making a public announcement, and then sells it the day after making the announcement.
- In the *Buy-and-hold* scheme, the official trader spends some fraction of his wealth the day before making a public announcement, and holds that position until the

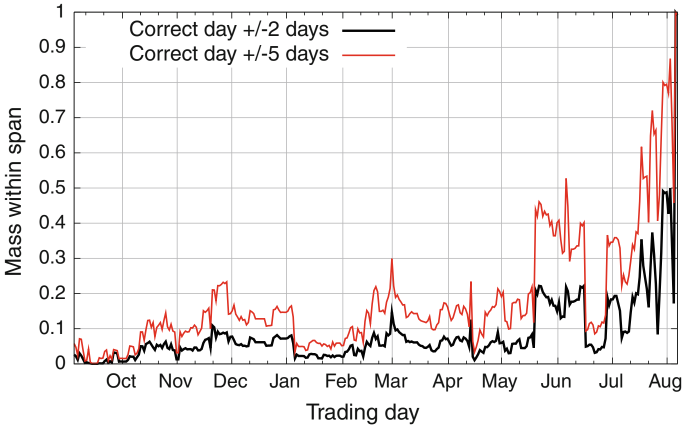

Fig. 6 The amount of probability mass around the correct opening day. The x-axis ranges over days the market was open. The *lines* indicate the mass of spans around the opening day on each trading day; *upper line* is for August 2 to 12, the *lower line* for August 5–9

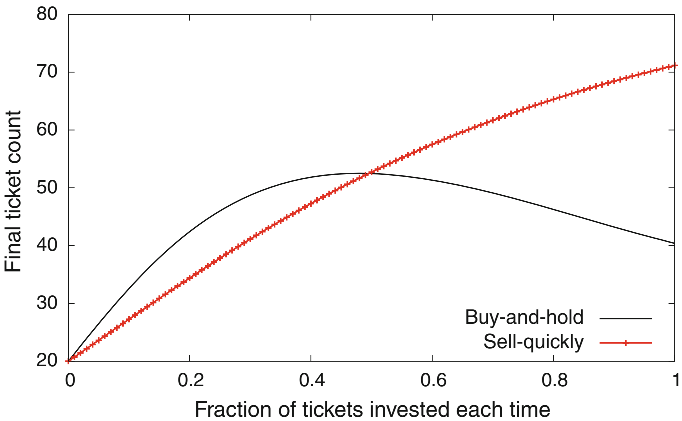

Fig. 7 The simulated final returns of a trader with inside information of official communications, starting with 20 tickets. The trader either holds their position until immediately before making a new communication ("Buy-and-hold") or closes their position the day after making an official communication ("Sell-quickly")

day before the next announcement is made (in the case of the final announcement, we assume he holds the position until the building opens).

There is some ambiguity inherent in the official statements. We interpreted "July" as meaning the whole month of July, and we interpreted "Approximately August 10th" as meaning the range August 7–13th, a 1 week range centered on August 10th. Observe that this span includes the correct opening day, August 7th.

The returns from both the Buy-and-hold and the Sell-quickly schemes depend on the fraction of wealth invested at each announcement. Figure 7 shows the final wealth of the inside-information trader for each scheme. Both schemes finish with more than

Table 3 Tickets and percent returns from following the two strategies after each announcement, where each agent invests all of his wealth in every action

Communication date	Sell-quickly	% Return	Buy-and-hold	% Return
October 15, 2008	23.83	19	55.40	177
February 14, 2009	43.33	82	81.80	48
July 23, 2009	68.17	57	21.80	−73
July 28, 2009	71.16	4	40.39	85

their 20 initially allotted tickets as long as the fraction of wealth invested with each announcement is positive.

Our simulation suggests that the inside-information trader would have profited from his better information by participating in the GHPM. Qualitatively, the official trader out-performed the market. As a counterfactual, however, our simulation could be wrong quantitatively. Specifically, our simulation is probably optimistically biased towards the returns of the official communications trader. Had the official trader actually participated in the markets, he would have driven up the prices of his desired spans, and so other traders probably would have been less inclined to drive up the prices of those spans even more. However, it is also possible trend-following traders might have driven the prices on those spans up even higher. This would result in higher quantitative returns for the inside-information trader.

Looking at the returns made by the schemes before each announcement helps us contextualize how much each announcement was anticipated. Table 3 displays this information, both in terms of total tickets after each action and percent return. The data in the table comes from each trader investing all of their wealth before each announcement, and so the final row of Table 3 corresponds to the rightmost values in Fig. 7.

Positions in the Sell-quickly scheme are liquidated a day after the official announcement is made. Therefore, the returns from the Sell-quickly scheme indicate how much the market moved in the short-term in response to the official information. Positions in the Buy-and-hold scheme are held until immediately before the next public communication. As a result, the returns from the Buy-and-hold scheme indicate how valuable each announcement was over the longer term.

Several of the values in the table merit further discussion.

- The highest-value trade in Table 3 is holding the "July" position from October 14th to February 13th. This produced a return of 177 % for the Buy-and-hold scheme, nearly tripling the Buy-and-hold trader's tickets. We attribute the success of this position to the announcement being nearly correct at a very early stage of the market.
- The highest return from the Sell-quickly strategy was the 82 % return from buying "August 3rd" on February 13th and selling it on February 15th. This exceptionally large return suggests that the market reacted quickly and dramatically to the announcement on February 14th.
- All the values in Table 3 are positive with the exception of the Buy-and-hold strategy buying "August 3rd" on July 22nd and selling the position on July 27th. In just 5 days, this position loses almost three-quarters of its value. This result,

combined with the very modest returns of only 4 % from the Sell-quickly trader's last action, suggests that the market anticipated the building would be delayed beyond August 3rd. However, the strongly positive returns (85 %) for the last action of the Buy-and-hold trader suggests that the market was anticipating a much longer delay than actually occurred. So, even though the probability of the building opening on August 3rd had fallen in the 5 days between announcements, the probability mass did not shift to the correct day, August 7th, but rather to later in August. This is confirmed by observing the skew of the probability distribution in Fig. 5 at the end of the market.

Both the Sell-quickly and the Buy-and-hold strategies produced positive earnings, which argues against the market fully anticipating every official communication. However, the losses of the Buy-and-hold strategy in the 5 days between denying and confirming a delayed opening suggest that the market correctly anticipated that the building would be delayed. However, the market appeared to anticipate a significantly longer delay than actually occurred.

5.2 Self-declared savviness

Recall from Fig. 1 that when traders signed up, they were asked "How savvy do you think you are relative to the average market participant?". They were given five choices, "Much less savvy", "Less savvy", "About the same" (the default selection), "More savvy", and "Much more savvy". Participants were informed that their answer to this question would not impact their payouts or the way they interacted with the market.

Because people are usually over-confident in various settings—and in prediction markets in particular (Forsythe et al. 1999; Graefe and Armstrong 2008)—it was our expectation that traders would be over-confident in their own abilities relative to others. Instead, we found the opposite.

5.2.1 Reported under-confidence

Based on prior studies of over-confidence in markets, we would expect to see most traders rate themselves as at least comparable to the average trader in the market. Table 4 shows our survey results. 77 traders described themselves as less or much less savvy than average, while only 13 traders described themselves as more savvy than average. Surprisingly, not a single trader listed themselves as much more savvy than the average trader.

Table 4 Self-assessment of savviness

Self-declared savviness	Number of traders
Much less than average	30 (17.8 %)
Less than average	47 (27.8 %)
Average	79 (46.7 %)
More than average	13 (7.7 %)
Much more than average	0

Table 5 Traders who self-identified as "more savvy than the average participant" in the market had dramatically lower median performance than other traders, while those traders identifying as "much less savvy than the average participant" had the same overall median performance as the trading population as a whole

Self-declared savviness	Median tickets
Much less than average	17.46
Less than average	16.78
Average	18.36
More than average	6.05
Much more than average	N/A

Why did we find traders under-confident, instead of over-confident, in their own abilities? Recent research by Moore and Healy (2008) on confidence sheds some light on this issue. They find that

> On difficult tasks, people . . . mistakenly believe that they are worse than others; on easy tasks, people . . . mistakenly believe they are better than others.

A novel market setting, such as the web-based automated market maker with span-based elicitation we used in the GHPM, is unfamiliar enough to a new trader as to seem potentially difficult. Prior market studies, because they have used traditional market interfaces that even the most casual participant is familiar with, would seem potentially less difficult and therefore would be susceptible to overconfidence.

5.2.2 Traders poorly predicted their own performance

We found that traders' self-reported savviness relative to other traders had little bearing on their relative performance. Table 5 groups traders by self-reported savviness and displays the group medians. The median over all traders was 17.46 tickets, identical to the least-savvy group and within a ticket of the two next-savvy groups. Ironically, traders identifying themselves as more savvy than the average trader had a median return more than 10 tickets lower than any other group.

5.3 On the two-ticket bonus

Recall that traders received a two-ticket bonus each week that they placed and held a bet in the market. Every Sunday night a script would examine a trader's holdings and see if there was an open bet created in the last week. If so, that trader would receive two extra tickets.

This suggests the question of how the bonus might have affected traders' strategies. The question is illuminated by the following observation. One simple strategy to exploit this bonus would be to place a bet on the entire span, once a week, using the two tickets gained in the previous week. This strategy would have ended with 116 tickets (20 initial tickets plus 2 bonus tickets each of 48 weeks), a final result that would have placed the strategy ninth among the trading population. Did any traders employ a similar strategy to ascend the leaderboard?

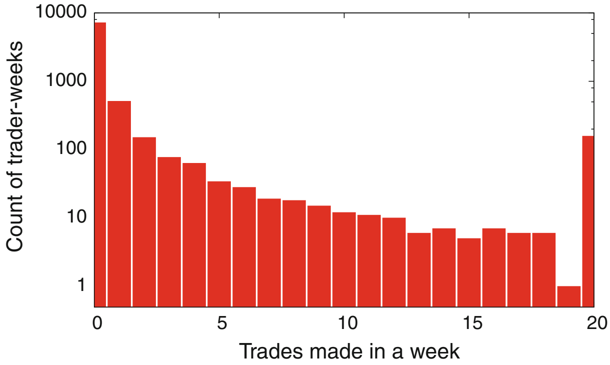

Fig. 8 The number of bets each trader placed each week in the market. The rightmost value on the plot, 20, includes all trader-weeks with a count of at least 20. The y-axis is log-scaled

Table 6 How tickets entered the trading market

Source	Number of tickets	Percent of tickets (%)
Initial endowment	3,380	58
Weekly bonus	2,278	39
Market maker subsidy	190	3

As a check, we manually investigated the top 15 traders in the market. We did not observe any obvious patterns of making putatively safe bets each week; most of these traders traded several times a week or not at all, and invested the majority of their tickets in the market in speculative bets rather than certain wagers or free tickets.

A more quantitative approach is demonstrated in Fig. 8. For each week and each trader, the plot shows the count of trader-weeks where the trader made exactly that many trades. (The rightmost column is a catch-all for 20 or more trades.) For instance, 10 trader-weeks saw a trader making exactly 12 trades in a week.

If trading behavior was skewed by the two-ticket bonus, we may expect to see the distribution of trader-weeks skewed towards a count of one, as traders would presumably recognize the opportunity for gains afforded by trading once each week, rather than not at all. However, Fig. 8 does not appear to show any obvious skew towards traders making exactly one trade a week.

Consequently, we were unable to find evidence that the two-ticket bonus substantially affected trader behavior. Of course, this does not imply two-ticket bonus had no effect on the market as a whole; the bonus may have still motivated traders to come to the website and participate more frequently than otherwise.

The two-ticket bonus was a vehicle by which a great deal of liquidity entered the market. Table 6 shows the source of tickets in the market.

Recall from Sect. 3.2 that the market maker's liquidity was under-provisioned, leading to price spikes. By putting the relative liquidity numbers in context, we see that the market maker's subsidy provided just 3 % of tickets entering the market. We suggest to future practitioners that more carefully balancing sources of liquidity can

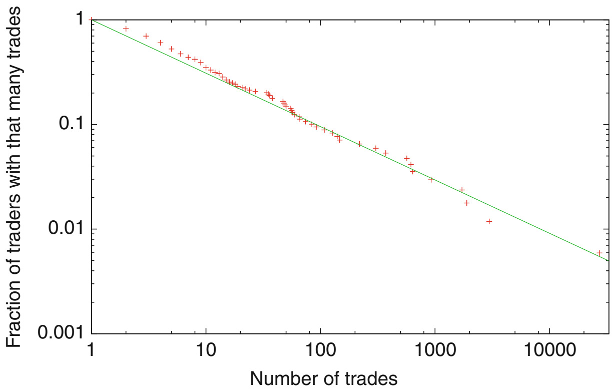

Fig. 9 A log - log plot showing the relationship between traders and the number of trades they placed. The straight line shows a power law fit for $\alpha = 0.51$

provide a guide for how to set a fixed level of liquidity (for liquidity-insensitive market makers) or for how to initially seed a variable level of liquidity (for liquidity-sensitive market makers).

5.4 Trade frequencies suggest a power law

The numbers of bets made by traders appear to closely fit a power law distribution. Figure 9 shows the relationship in terms of the probability of a trader having more than a certain number of trades from our data set, and the best-fitting power law distribution. (We also tried a log-normal distribution and the fit was poor.)

Unfortunately, with only 169 traders we can not assert an appropriate level of statistical significance, so we cannot rule out the data being generated by other distributions. However on a log–log plot the data do appear to snugly fit the canonical straight line of a power law distribution over several orders of magnitude.

5.5 Trading by a bot

Conventionally, when we think about prediction markets, we think about a collection of individuals making probability judgments. This is a quality distinct from traditional exchanges, in which automated trading is common and frequent. But as Berg et al. (2001) discuss, trading bots make up a large fraction of the observed volume in the Iowa Electronic Markets (IEM), the longest-running prediction market, and must be considered in any sort of qualitative summary of the properties of prediction markets. We found that trade in the GHPM was also dominated by a bot.

This was surprising because we did not make automated trading easy. The GHPM did not use an API, so any trading bot would have to come up with a way to parse the web page and simulate its user's actions on the page. Jim, a graduate student in the Computer Science Department, took two days to write a trading bot. The bot fit the current prices to a mixture of Gaussian distributions and identified trading opportunities based on deviations from the fit.

The bot made 68.5 % of the trades in the market (27,311 of 39,842). The median number of trades placed for all traders was five.

Jim's bot did well in the market; at its peak it was the second-highest-valued trader. Jim turned his bot off after the e-mail of February 14th and began trading manually. He ended up losing the bulk of his tickets by betting on the building opening earlier than it actually did, finishing 158th of the 210 registered users and 117th out of the 169 traders.

5.6 Trader-level data supports the Marginal Trader Hypothesis

Prediction markets have been demonstrated to be at least as accurate as, and in many cases more accurate than, predictive techniques like polls (Berg et al. 2001; Goel et al. 2010). How do markets incorporate information and generate good prices? A line of research from the administrators of the IEM has suggested that a small group of so-called *marginal traders* are responsible for producing accurate results in prediction markets (Forsythe et al. 1992, 1999; Berg et al. 2001; Oliven and Rietz 2004). This group, which appears to be about 10 % of the traders in the IEM, essentially arbitrage the remainder of the market participants. This theory of how markets function is called the *Marginal Trader Hypothesis*.

In this section, we show that the GHPM supports the Marginal Trader Hypothesis. To our knowledge this is the first confirmation of the Marginal Trader Hypothesis within a market that used an automated market maker. This finding is significant because using an automated market maker prevents less sophisticated agents from making many of the errors that could be thought to drive the lopsided distribution of performance that the Marginal Trader Hypothesis predicts. We also argue that the GHPM does not support one of the most intriguing findings of the literature: that marginal traders are disproportionately male.

5.6.1 Why isn't it easy to identify marginal traders?

The techniques from the IEM literature do not apply to our market setting. In Forsythe et al. (1992) marginal traders were identified by the particular kinds of trades they placed. Specifically, marginal traders were those traders that provided liquidity and set prices by placing limit orders (orders with a limiting price, e.g., "I will buy 10 contracts at a price-per-contract of no greater than 60 cents"), rather than market orders ("I will buy 10 contracts"). Put another way, marginal traders served as market makers. In later work, this connection was made more explicit. Oliven and Rietz (2004) classify traders as either marginal, price-setting, market makers, or as non-marginal price takers.

One of the key findings of this literature is that non-marginal traders violated the *law of one price*. To illustrate what the law of one price is, consider the example of the Red Sox and Yankees playing a baseball game with a traditional prediction market not equipped with an automated market maker. Imagine that the current order books on the events are given in Table 7.

Now consider two actions at these prices: buying a share of Red Sox stock, or selling a share of Yankees stock. If a trader buys a share of Red Sox stock, his payoffs are

Table 7 Hypothetical prices in a baseball game prediction market

Contract	Best bid	Best ask
Red Sox	0.34	0.40
Yankees	0.63	0.65

$\{0.6, -0.4\}$, where ω_1 = Red Sox win and ω_2 = Yankees win. If a trader sells a share of Yankees stock, his payoffs are $\{0.63, -0.37\}$. Observe that this payoff vector strictly dominates the payoff from buying a share of Red Sox stock—no matter whether the Red Sox or Yankees win the game, the trader receives a higher payoff from the latter action. Consequently, if a trader were to do the former action and buy a share of Red Sox stock, he is said to be violating the law of one price. More formally, violations of the law of one price occur when an agent takes on a strictly-worse payoff vector than one that could be constructed at the current market prices. Interestingly, even though violating the law of one price is a dominated action, the IEM literature has found that non-marginal traders perform these actions frequently (Oliven and Rietz 2004).

In summary, the literature from the IEM gives us two trade-level ways of distinguishing between marginal and non-marginal traders: marginal traders place limit orders, and non-marginal traders violate the law of one price. Unfortunately, when using the LMSR, neither of these distinguishing characteristics are available to us. First, the LMSR *only* uses market orders, so agents cannot place limit orders. Therefore, traders cannot be price setters, only price takers. Second, the LMSR precludes any violation of the law of one price, because the LMSR maintains a probability distribution over the different states of the world. This probability distribution enforces the condition that equivalent bets for the set of states "A" and against the set of states "not A" have equivalent payoff vectors.

An example will be helpful in seeing why this is the case. Consider our example of the Red Sox-Yankees baseball game again. "Buying the Red Sox" is equivalent to taking on the contract $\{1, 0\}$. "Selling the Yankees" is equivalent to taking on the contract $\{0, -1\}$. For the payoff vector associated with the former contract, we subtract the amount we must pay for the contract from the vector $\{1, 0\}$. For the payoff vector associated with the latter contract, we add the money we receive from selling the contract to the vector $\{0, -1\}$. We will show that for any initial x and any arbitrage-free cost function C (in the sense of Sect. 3.2), a set that includes the LMSR, the payoff vector formed by these two bets is equivalent.

The two bets are equivalent if

$$\{1, 0\} - (C(x + \{1, 0\}) - C(x))\,\mathbb{1} = -(C(x + \{0, -1\}) - C(x))\,\mathbb{1} + \{0, -1\}$$

By re-arranging, we see that this condition is

$$\begin{aligned}
(C(x + \{1, 0\}) - C(x)) - (C(x + \{0, -1\}) - C(x)) &= 1\\
C(x + \{1, 0\}) - C(x + \{0, -1\}) &= 1\\
C(x + \{1, 0\}) &= C(x + \{0, -1\}) + 1\\
C(x + \{1, 0\}) &= C(x + \{0, -1\} + \{1, 1\})\\
C(x + \{1, 0\}) &= C(x + \{1, 0\})
\end{aligned}$$

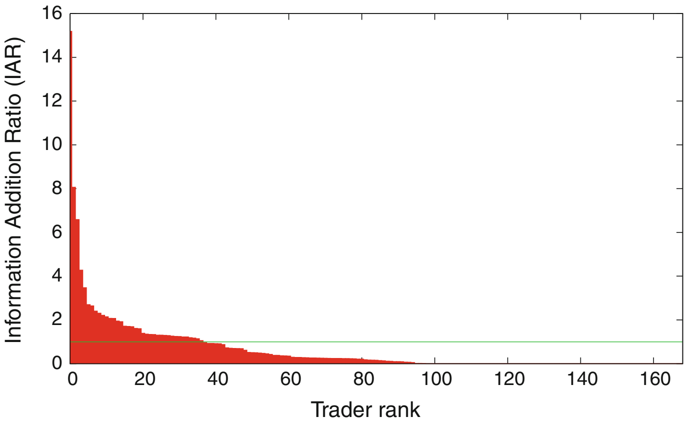

Fig. 10 The distribution of IAR of traders ordered by rank. The *straight line* shows an IAR of 1 (one ticket expected per ticket wagered)

Here the penultimate step relies on the arbitrage-freeness of the cost function (i.e., $C(x) + 1 = C(x + \mathbf{1})$).

Given that we cannot identify marginal traders in the ways suggested by prior literature, in the next section we discuss how we employed the spirit of the original work of the IEM to select a candidate set of marginal traders.

5.6.2 Identifying marginal traders in the GHPM

The IEM researchers made their criteria for selected marginal traders on the basis of informative trading patterns. They then showed that their set of marginal traders had better performance and more frequent trade than their set of non-marginal traders. Here, we take a similar approach. We start by describing a way of gauging the information content of the bets a trader makes, and how this metric produces an intuitive way to isolate a set of marginal traders.

We dub the metric we used to assess the information content of a trader's bets the *Information Addition Ratio (IAR)*. This measure attempts to separate a trader's return from speculative activities from a trader's return from information-adding activities. It answers the question "If we see a trader making a one-ticket bet, what is her expected return if she were to hold that bet until the market closes?". A return of one ticket on each ticket invested is always available to a trader by betting on the entire range of exhaustive contracts. Traders who inject valuable information into the market will have an IAR greater than one, while traders who have a deleterious impact on information will have an IAR of less than one. Essentially, IAR measures how much each trader increased the price of August 7th, the correct opening day. IAR is an attempt to compress a complex concept into a scalar, and such an enormous dimensionality reduction is inherently lossy. IAR places a focus exclusively on rewarding traders for making bets that raised the price of the correct opening day.

Figure 10 displays the distribution of IARs. It is heavily skewed and unequal. The median trader had a return of 0.16 tickets per ticket bet, and 79 traders (47 %)

Table 8 Gini coefficients are a standard measure of the degree of inequality of a distribution

Data set	Gini coefficient
Normal distribution $\mu = 5, \sigma = 1$	0.113
Denmark income	0.247
Uniform distribution	0.333
United states income	0.408
Log-normal distribution $\mu = 5, \sigma = 1$	0.521
GHPM tickets	0.700
Namibia income	0.743
GHPM IARs	0.762

As this table shows, the distribution of both information (IARs) as well as overall performance (tickets) were extremely unequal. For reference, we include country income inequality coefficients from the Watkins (2008); Denmark had the lowest coefficient and Namibia the highest

placed all their bets on losing intervals, i.e., spans that did not include the correct date, August 7th. This is surprising for several reasons. First, a return of one ticket per ticket bet was *always* available to traders by betting on the entire span. Second, in the ternary elicitation interface, one of the bets offered will always include August 7th, and the other will not, so there was no inherent bias against traders making correct bets. Finally, the median number of bets per trader was five, meaning that the bulk of traders made poor judgements several times, not just once.

We can gauge how skewed a distribution is at a glance by measuring its Gini coefficient, a standard measure of inequality. Assuming we have the data points $x_1 \leq x_2 \leq \cdots \leq x_n$, the Gini coefficient, G, of the sample is given by

$$G = \frac{2 \sum_i i x_i}{n \sum_i x_i} - \frac{n+1}{n}$$

The Gini coefficient ranges from zero to one and can be thought of as a measure of how unequal drawn samples are, with particular sensitivity to large outliers. Table 8 displays the Gini coefficients for the GHPM in context with other distributions. The Gini coefficients of both final tickets and IARs is very high, which indicates the underlying distributions are skewed. Taken as a whole, these results indicate that the majority of market participants consistently made judgments that hurt the accuracy of the GHPM.

On the other side, there was clearly a small and select group of traders responsible for actually making the GHPM produce meaningful prices. Only 37 traders (22 %) had an IAR of more than one, and only 13 traders (8 %) had an IAR of more than two. A trader with an IAR of greater than two was making sophisticated judgements to bet correctly on less than half of the market's probability mass. A trader could achieve an IAR of more than one by making not-very-nuanced bets with the market maker (e.g., by betting on all but the earliest day), but there is no way a trader could have an IAR of more than two without making nuanced judgements. Consequently, we advance these 13 traders as our candidate set of marginal traders.

Table 9 The performance of marginal and non-marginal traders in the 1988 IEM presidential market (Forsythe et al. 1992))

	IEM marginal	IEM non-marginal
Fraction of traders	0.11	0.89
Median percent return	9.6	0.0
Average total investment	56	21

Table 10 The relative performance of our candidate set of marginal traders in the GHPM

	GHPM marginal	GHPM non-marginal
Fraction of traders	0.08	0.92
Median tickets	79.2	16.3
Median trades	147	5

5.6.3 Comparison of marginal and non-marginal traders

In this section, we will compare the group we selected as candidate marginal traders to the group we selected as non-marginal traders. We show that, just like in the IEM literature, the performance and level of involvement of the marginal traders was much higher than the non-marginal traders. Thus, we argue that the GHPM supported the Marginal Trader Hypothesis.

Table 9, taken from Forsythe et al. (1992), displays the performance of traders in the 1988 IEM presidential election market. Their set of marginal traders were about ten percent of participants, and the marginal traders had a much higher percent return and invested much more in the market than the non-marginal traders.

Table 10 displays the performance of our candidate set of marginal traders relative to non-marginal traders in the GHPM. Just like in the IEM, about 10 % of traders are classified as marginal. Additionally, the marginal traders in the GHPM were a much more active presence than non-marginal traders. The median marginal trader made almost 30 times more bets than the median non-marginal trader.

Finally, just like in the IEM, the performance of the marginal traders was much better than the performance of the non-marginal traders. However, whereas the median marginal trader in the IEM had a rate of return of 9.6 %, the median marginal trader in the GHPM had a rate of return of almost 300 %. We attribute this discrepancy to the weekly ticket handout, as well as the subsidy given up by the market maker (recall that the LMSR runs at a loss). We conjecture that had the IEM awarded traders for participation and had the IEM used a market making agent that subsidized traders, then most of these rewards would have gone to the marginal traders because they were the most active participants.

One possible criticism is that the results of this section were a foregone conclusion, and that our criteria for classifying a trader as marginal necessitated that they end up being an active, profitable participant. But just like in Forsythe et al. (1992), we used trade-level attributes to decide whether or not a trader would be classified as marginal, and there was no guarantee that a trader selected as marginal would end up with the most tickets or the most trades. It is easy to imagine, in fact, traders who could qualify as marginal by our definition that would not be active or involved participants; such

Table 11 The gender composition of the marginal trader pools in our study and in the IEM (as reported in Forsythe et al. (1992))

	IEM marginal	IEM non-marginal	GHPM marginal	GHPM non-marginal
Number	22	170	13	156
Fraction male	1.00	0.68	0.77	0.74

a trader could merely have placed a single small bet on the market opening in July, August, or September soon after the market's initiation. This would have produced an IAR of greater than two for that trader without significant involvement in the market and without that trader amassing a large number of tickets. Furthermore, IAR need not have any correspondence to a trader's actual returns from the bets they placed; if they were to sell their correct bets before the market's expiration, they would earn fewer tickets or, depending on short-term prevailing market prices, could even lose tickets on those bets.

However, our results show that this was not the case, and that traders whose bets raised the price of the correct opening day tended to be by far the most profitable and active traders in the GHPM as a whole. In terms of tickets, our set of marginal traders included the top trader, four of the top five, and six of the top ten. In terms of trades, our set of marginal traders included the top trader, four of the top five, and seven of the top ten.

5.6.4 Are only men marginal traders?

One of the most curious findings of Forsythe et al. (1992) was that their pool of marginal traders was *exclusively male*. In contrast, our pool of marginal traders was not statistically significantly different in gender composition from our pool of traders as a whole. Table 11 compares the results of the two prediction markets. (These results are not an artifact of using IAR to select marginal traders, because women were also well-represented under alternative selection criteria. By final ticket count, three of the top ten traders were women. By number of trades, two of the top ten traders were women.)

For the null hypothesis that the GHPM marginal and non-marginal traders are drawn from the same gender distribution, we get a p-value of 0.60. Since $p > 0.05$, this means it is not statistically significant to reject the null hypothesis.

Consequently, our results do not support the hypothesis that the gender composition of the marginal traders differs from the gender composition of non-marginal traders. Since there does not seem to be any causal reason that women should be worse traders than men, and Forsythe et al. (1992) do not justify any mechanism for their gender findings, we contend that the hypothesis of women being less likely to be marginal traders should probably be dismissed, or at the least, merits further study before acceptance.

6 Conclusion

The Gates Hillman Prediction Market (GHPM) represented the largest test faced by automated market making in prediction markets. It was a long-lived market with hundreds of participants, hundreds of events, and tens of thousands of trades. By testing the boundaries of automated market making through actual implementation, we can get a better perspective on where new research should be directed. The GHPM uncovered two significant shortcomings in current market maker designs:

- Spikiness of prices of similar events at any snapshot in time. Unfortunately, we proved that every market maker meeting three simple properties will induce spikes. We also demonstrated a new interaction interface that enabled participation by both sophisticated and unsophisticated users over a very large event space. Perhaps other interaction interfaces could be developed that lead traders to place bets that might better reflect their beliefs while still being simple enough for unsophisticated users. One suggestion would be to have users wager on payout vectors, with the default being an equal payout on each day for simplicity, as in the GHPM.
- Liquidity insensitivity that led to price volatility even in mature stages of the market. As we showed, the tenet of no arbitrage (Chen and Pennock 2007) within the prediction market literature keeps liquidity at a constant level throughout a market's tenure. Future market makers should be liquidity sensitive: they should make prices "stiffer" (i.e., the price changes less as a function of the amount that is bet) in markets where lots of trade volume has been observed. The designs of such market makers are discussed in Othman et al. (2010) and Othman and Sandholm (2012).

The GHPM produced a rich data set of identity-linked trades, and our access to individual traders allowed us to interview them about the strategies they employed in the market. With this unique data set, we were able to provide in-depth study of the market's microstructure. By simulating the performance of a trader with inside information of official communications, we showed that the market reacted to official communications, but correctly predicted a delay in advance of official notification.

Using the data set, we argued that the GHPM supports the Marginal Trader Hypothesis. This hypothesis, originally derived and supported by the administrators of the Iowa Electronic Markets, argues that a small fraction of traders (about 10 %) are responsible for injecting good information in the market and essentially arbitrage the remainder of the trading population. Using trade-level data, we showed that the GHPM had a similarly skewed structure to the amount of information traders injected: a small fraction made discriminating, correct bets that raised the price of the correct opening day, while the median trader made five bets that all lowered the price of the correct opening day. While the IEM administrators found that the pool of marginal traders was disproportionately (in fact, exclusively) male, the data in the GHPM does not support rejecting the hypothesis that the pools of marginal and non-marginal traders have the same gender compositions. As we discussed, the automated market maker that mediated bets in the GHPM precluded less sophisticated traders from making mistakes like violating the law of one price. Conceivably, this could make the Marginal Trader Hypothesis no longer hold by protecting less savvy traders from being arbitraged. Consequently, it is significant that our study is the first to confirm the Marginal Trader Hypothesis in a market equipped with an automated market maker.

Acknowledgments We thank Don McGillen and Yahoo! for their generous lead sponsorship of the GHPM prizes. We also recognize Peter Lee, erstwhile CMU Computer Science Department head, as well as Chuck Schneider and Rick Zuchelli of the CMU Bookstore for their donations. We are also very thankful to David Pennock and Daniel Reeves, formerly of Yahoo! Research, for discussions in the GHPM's formative stages. Finally, we gratefully acknowledge the feedback and guidance of an anonymous journal reviewer as well as the stewardship of the editorial staff of the Review of Economic Design. This material is based upon work supported by the National Science Foundation under grants CCF-1101668, IIS-0905390, and IIS-0964579. Othman is supported by the Google PhD Fellowship in Market Algorithms.

References

Agrawal S, Delage E, Peters M, Wang Z, Ye Y (2009) A unified framework for dynamic pari-mutuel information market design. In: ACM conference on electronic commerce (EC). ACM, New York, pp 255–264

Ali M (1977) Probability and utility estimates for racetrack bettors. J Polit Econ 85(4):803

Ben-Tal A, Teboulle M (2007) An old-new concept of convex risk measures: the optimized certainty equivalent. Math Finance 17(3):449–476

Berg J, Forsythe R, Nelson F, Rietz T (2001) Results from a dozen years of election futures markets research. In: Handbook of experimental economic results, vol 1, pp 486–515

Chakraborty M, Das S, Lavoie A, Magdon-Ismail M, Naamad Y (2011) Instructor rating markets. In: Conference on auctions, market mechanisms and their applications (AMMA)

Chen Y, Pennock DM (2007) A utility framework for bounded-loss market makers. In: Proceedings of the 23rd annual conference on uncertainty in artificial intelligence (UAI), pp 49–56

Chen Y, Pennock DM (2010) Designing markets for prediction. AI Mag 31(4):41–52

Chen Y, Fortnow L, Lambert N, Pennock DM, Wortman J (2008) Complexity of combinatorial market makers. In: ACM conference on electronic commerce (EC). ACM, New York, pp 190–199

Cowgill B, Wolfers J, Zitzewitz E (2009) Using prediction markets to track information flows: evidence from Google. In: Conference on auctions, market mechanisms and their applications (AMMA)

Forsythe R, Nelson F, Neumann G, Wright J (1992) Anatomy of an experimental political stock market. Am Econ Rev 82(5):1142–1161

Forsythe R, Rietz T, Ross T (1999) Wishes, expectations and actions: a survey on price formation in election stock markets. J Econ Behav Organ 39(1):83–110

Fortnow L, Kilian J, Pennock DM, Wellman MP (2003) Betting boolean-style: a framework for trading in securities based on logical formulas. In: ACM conference on electronic commerce (EC). ACM, New York, pp 144–155

Goel S, Pennock D, Reeves D, Yu C (2008) Yoopick: a combinatorial sports prediction market. In: Proceedings of the national conference on artificial intelligence (AAAI), pp 1880–1881

Goel S, Reeves D, Watts D, Pennock D (2010) Prediction without markets. In: ACM conference on electronic commerce (EC). ACM, New York, pp 357–366

Graefe A, Armstrong JS (2008) Can you beat the market? Accuracy of individual and group post-prediction market judgments. In: Third workshop on prediction markets, part of the ACM conference on electronic commerce (EC)

Hanson R (2003) Combinatorial information market design. Inf Syst Front 5(1):107–119

Hanson R (2007) Logarithmic market scoring rules for modular combinatorial information aggregation. J Predict Mark 1(1):1–15

Healy P, Linardi S, Lowery J, Ledyard J (2010) Prediction markets: alternative mechanisms for complex environments with few traders. Manag Sci 56(11):1977–1996

Lambert NS, Pennock DM, Shoham Y (2008) Eliciting properties of probability distributions. In: Proceedings of the 9th ACM conference on electronic commerce (EC). ACM, New York

Ledyard J, Hanson R, Ishikida T (2009) An experimental test of combinatorial information markets. J Econ Behav Organ 69(2):182–189

Moore D, Healy P (2008) The trouble with overconfidence. Psychol Rev 115(2):502–517

Oliven K, Rietz T (2004) Suckers are born but markets are made: individual rationality, arbitrage, and market efficiency on an electronic futures market. Manag Sci 50(3):336–351

Ostrovsky M (2009) Information aggregation in dynamic markets with strategic traders. In: ACM Conference on Electronic Commerce (EC), pp 253–254

Othman A, Sandholm T (2010) When Do Markets with Simple Agents Fail? In: International conference on autonomous agents and multi-agent systems (AAMAS). Canada, Toronto, pp 865–872
Othman A, Sandholm T (2011) Liquidity-sensitive market makers via homogeneous risk measures. In: International workshop on internet and network economics (WINE)
Othman A, Sandholm T (2012) Profit-charging market makers with bounded loss, vanishing bid/ask spreads, and unlimited market depth.In: ACM conference on electronic commerce (EC), pp 790–807
Othman A, Pennock DM, Reeves DM, Sandholm T (2010) A practical liquidity-sensitive automated market maker. In: ACM conference on electronic commerce (EC), pp 377–386
Pennock D, Sami R (2007) Computational Aspects of Prediction Markets. In: Algorithmic game theory, chapter 26. Cambridge University Press, Cambridge, pp 651–674
Watkins K (2008) Human Development Report 2007/2008. United Nations Development Programme, New York
Wolfers J, Zitzewitz E (2004) Prediction markets. J Econ Perspect 18(2):107–126
Wolfers J, Zitzewitz E (2006) Five open questions about prediction markets. Technical report 1975, institute for the study of labor (IZA), Feb 2006