

Incentivizing Optimal Forecasting Groups in Prediction Markets

NICHOLAS VON TURKOVICH

Duke University
nbv3@duke.edu

April 26, 2017

Abstract

I. INTRODUCTION

Wisdom of crowds describes a phenomenon in which a group of individuals perform a task or make a decision with an acumen greater than any single individual. Oftentimes, the disparity between a single individual performing a task and the collective is stark. Possibly the most quoted and most widely referenced single instance of wisdom of the crowds is attributed to Sir Francis Galton’s observations of a crowd estimating the weight of an ox.¹ He noted that the crowd’s average prediction was within 1 percent of the oxen’s true weight. Since then, wisdom of the crowds has garnered interest from various industries and academic fields as more individuals try to ascertain what makes a crowd intelligent and how crowd intelligence be utilized to perform tasks that would be difficult for any single individual. The most prevalent example of crowds being used to accomplish tasks comes in the form of markets, where groups of individuals contribute their private information and perspective for personal gain, and in so doing, help to attain an accurate estimation of a quantity. This quantity could be the value of a particular company’s stock or the likelihood of some event happening in the future.

A prediction market acts as a vehicle for obtaining likelihood estimates of a particular event or set of events occurring.² Participants in the market pay for a contract that is tied to a particular probability distribution over the set of events in question. The market determines both how to aggregate information elicited from market participants as well as how to reward individual contracts according to some payoff function. There is strong evidence

backing up the accuracy of different prediction markets [2].

There is also substantial research focused on crowd dynamics and what characteristics make for an optimal crowd in a setting where the objective is to estimate a given quantity [3]. The purpose of this paper is to determine whether or not we can build a prediction market that changes its incentive scheme to ensure that the participating population closely matches the optimal composition given the pool of potential participants. We will first provide a brief background on scoring rules for prediction markets. Then, we will take a more in-depth look at some current literature regarding composition of forecasting groups [3]. This background will allow us to frame the motivating question for this research. After proposing the incentive structure that is meant to ensure optimal group composition, we will then evaluate the benefits and shortcomings of the scheme and suggest both improvements and directions for future work.

II. BACKGROUND

i. Scoring Rules

In a prediction market setting, participants are asked to report a vector of probabilities denoting the likelihoods of different events occurring. The reward for a particular vector of probabilities is a function of the vector and the set of events that eventually occurred. This function is known as a scoring rule and is typically denoted as:

$$S(\vec{p}, \omega) \tag{1}$$

where \vec{p} is the vector of probabilities bid by the market participant and ω is the actual event that occurred. Note that in predicting a single event, the vector \vec{p} has length 1.

¹His original paper, *Vox Populi*, was published in *Nature* in March of 1907. doi:10.1038/075450a0

²See [1] for more details.

There are certain properties of scoring rules that elicit different behavior from rational agents. A scoring rule is called proper if the agent cannot improve their expected payoff by misreporting their belief or what they believe to be the true probabilities over events. A scoring rule is strictly proper if the set of probability vectors that maximize an agent's expected value has a cardinality of 1, where the single element of the set is their true belief. Therefore, for both types of scoring rules, there is no incentive to report some $\hat{p} \neq \vec{p}$ and for a strictly proper scoring rule, the agent can maximize their expected payment only by reporting \vec{p} .

The expected value of an agent predicting a the probability of an event is given by:

$$G(\vec{p}) = pS(p, 1) + (1 - p)S(p, 0) \quad (2)$$

We know that for a (strictly) proper scoring rule, the function $G(\vec{p})$ is (strictly) convex. Gneiting and Raftery [4] generalize the conditions that must be satisfied by the function $G(\vec{p})$ in order for a scoring rule to be proper or strictly proper.

ii. Optimal Forecasting Groups

The following model and results stem from the work by Lamberson and Page (LP) [3] as they characterized what makes an optimal forecasting group under their assumptions. This section summarizes the details of their model and results that are relevant to the work in this paper.

ii.1 Basic Framework

In the model, agents have a goal of estimating some quantity V . Each of the M agents is characterized by a random variable s_i , which is their prediction of the quantity V . The random variable $\epsilon_i = s_i - V$ is the error in prediction for the i_{th} agent. The aggregate estimate of the crowd of agents is $G(\vec{s}) = \frac{1}{M} \sum_i s_i$. In [3], LP focuses solely on the simple average as a way of aggregating individual estimates. Finally, the metric for accuracy in LP is group squared error, $(G(\vec{s}) - V)^2$. LP then uses the *bias-variance-covariance* decomposition to expand the expected value of the group squared error:

$$E[(G(\vec{s}) - V)^2] = \overline{bias}(\vec{s})^2 + \frac{1}{M} \overline{var}(\vec{s}) + (1 - \frac{1}{M}) \overline{cov}(\vec{s}) \quad (3)$$

where the average bias of signals is given by:

$$\overline{bias}(\vec{s}) = \frac{1}{M} \sum_{i=1}^M (E[\epsilon_{s_i}]) \quad (4)$$

the average error variance is given by:

$$\overline{var}(\vec{s}) = \frac{1}{M} \sum_{i=1}^M var[\epsilon_{s_i}] \quad (5)$$

and the average error covariance is given by:

$$\overline{cov}(\vec{s}) = \frac{1}{M(M-1)} \sum_{i=1}^M \sum_{j \neq i} cov(\epsilon_{s_i}, \epsilon_{s_j}) \quad (6)$$

ii.2 Group Composition

Now, let there be two types of forecasters in the crowd characterized by prediction random variables a and b . Furthermore, LP assumes the agents to be unbiased such that:

$$E[\epsilon_a] = 0 \rightarrow E[\epsilon_a^2] = var(\epsilon_a) - (E[\epsilon_a])^2 = var(\epsilon_a) \quad (7)$$

$$E[\epsilon_b] = 0 \rightarrow E[\epsilon_b^2] = var(\epsilon_b) - (E[\epsilon_b])^2 = var(\epsilon_b) \quad (8)$$

LP goes on to define $cov(\epsilon_a)$ as the covariance in prediction errors between two individual agents of type a (i.e. intra-group covariance). Likewise, $cov(\epsilon_b)$ denotes the covariance in prediction errors between two agents of type b . Finally, $cov(\epsilon_a, \epsilon_b)$ (inter-group covariance) is the covariance in prediction errors between the two types. In this setting, covariance indicates how similarly prediction errors for agents will vary together. For example, errors in prediction may covary if the agents draw upon similar resources or information in making their prediction or if they have similar perspectives on the problem. The final relevant assumption that LP makes is the enforcement of a condition, *type coherence*.

Type coherence is satisfied if the quantity:

$$TC(a, b) = cov(\epsilon_a) + cov(\epsilon_b) - 2cov(\epsilon_a, \epsilon_b) \quad (9)$$

is strictly positive: $TC(a, b) > 0$. This assumption is fairly reasonable as it states for the results of their analysis to hold, the types' errors must covary more within the

two groups than they do between the two groups. In other words, the errors in prediction amongst members of groups should be more related than the errors across groups.

ii.3 Relevant Results from LP

From equation 3, LP gives the following for the expected squared error in average prediction for a group with types a and b :

$$E[(G(s) - V)^2] = \frac{A\text{var}(\epsilon_a) + B\text{var}(\epsilon_b) + 2\left(\frac{A}{2}\right)\text{cov}(\epsilon_a) + 2\left(\frac{B}{2}\right)\text{cov}(\epsilon_b) + 2AB\text{cov}(\epsilon_a, \epsilon_b)}{M^2}$$

where $M = A + B$. Here A indicates the quantity of type a in the crowd and B indicates the quantity of type b in the crowd. Recall that the agents are still assumed to be unbiased.

LP then treats the above equation as a function of A and finds that the expected group error is minimized when the proportion of type a agents in the crowd is:

$$\frac{A}{M} = \frac{[\text{var}(\epsilon_b) - \text{var}(\epsilon_a)] - [\text{cov}(\epsilon_b) - \text{cov}(\epsilon_a)]}{2M \cdot \text{TC}(a, b)} + \frac{\text{cov}(\epsilon_b) - \text{cov}(\epsilon_a, \epsilon_b)}{\text{TC}(a, b)}$$

From this equation, we can see that as $M \rightarrow \infty$, the optimal fraction of type a individuals approaches $\frac{\text{cov}(\epsilon_b) - \text{cov}(\epsilon_a, \epsilon_b)}{\text{TC}(a, b)}$. This second equation shows that the proportion of a type should increase as the variance (i.e. inaccuracy) of type a errors decreases with respect to the variance of type b errors. Furthermore, when we take the limit as $M \rightarrow \infty$, accuracy becomes less of a determining factor in the proportions of the two types. As the intra-group covariance of one group increases (the errors of that group covary more), the proportion of the opposing group needed to minimize the expected group error increases. Hence, when as the group gets larger, it favors the type with more internal disagreement in prediction errors.

This concludes the summary of the results of LP, which serve as a stepping stone for the continued work proposed in this paper.

III. INCENTIVE SCHEME FOR MAINTAINING OPTIMAL GROUP COMPOSITION

The insights from LP [3] regarding optimal composition of forecasting groups can be extended in a number of different ways. In the context of prediction markets though, we can try to use this knowledge to encourage types that would contribute favorably to the group's

average error and discourage the type that would contribute unfavorably to the group's average error. Using additional incentives or disincentives in addition to the use of proper scoring rules, we can encourage agents to participate in the market if they have valuable information regarding the event being predicted and if they are of a type that would improve the group's error in expectation.

We assume that the agents know their type and are aware of the current group composition.

i. Strict Incentive Scheme

Ideally, the following should hold in the incentive scheme for an individual deciding whether or not to participate in the prediction market. Here π denotes the utility or expected reward for participation.

$$\pi \propto \text{accuracy of prediction} \quad (10)$$

$$\pi \propto \text{value of type to group} \quad (11)$$

For brevity, let the expected squared error of the group prediction from [3], $E[(G(\vec{s}) - V)^2]$, be denoted as $GE(A, B)$. In a traditional prediction market setting, where types do not influence the incentives for participation, the expected utility, π would simply be:

$$\pi(\vec{p}) = G(\vec{p}) \quad (12)$$

where $G(\vec{p})$ is the expected reward for reporting \vec{p} . Now, to incentivize types from participating or refraining to participate, the following is the utility function with the addition of another term.

$$\pi(\vec{p}, \theta) = G(\vec{p}) + C[GE(A, B) - GE(A + 1, B)] \text{ if } \theta = a \quad (13)$$

$$\pi(\vec{p}, \theta) = G(\vec{p}) + C[GE(A, B) - GE(A, B + 1)] \text{ if } \theta = b \quad (14)$$

where C is a constant chosen by the market organizer and $GE(A, B)$ denotes the current proportions of agents in the crowd. Without asking for an individual to report their type, C must be chosen in such a way that the type that is less favored or will produce more error in the group estimation of the probability is completely disincentivized

from participating. This means that for the unfavored type, $C[GE(A, B) - GE(A + 1, B)] \leq -|\max_{\vec{p}} G(\vec{p})|$. This would ensure that for the highest possible $G(\vec{p})$ value for the unwanted type, that type would have no incentive to bid. For the majority of the paper, we will consider how we choose C in the context of a quadratic scoring rule where the maximum possible expected payoff is 1.

Both $GE(A, B) - GE(A + 1, B)$ and $GE(A, B) - GE(A, B + 1)$ can take on positive or negative values.

Case 1: $GE(A, B) - GE(A + 1, B) < 0$,
 $GE(A, B) - GE(A, B + 1) < 0$

In this situation, the addition of a single agent of either type will be detrimental for the group error in expectation. This could happen in a situation where the current proportions are close to optimal for some crowd size M . However, as the market organizer, we don't necessarily want to halt all participation. Recall the first result from LP. The expected square error is inversely proportional to M^2 . Therefore, as the crowd gets larger, overall error shrinks. Therefore, C should be picked such that the agent that would contribute the least to the expected error is included:

$$C = \min\left\{\frac{1}{|GE(A, B) - GE(A + 1, B)|}, \frac{1}{|GE(A, B) - GE(A, B + 1)|}\right\}$$

Case 2: $GE(A, B) - GE(A + 1, B) > 0$,
 $GE(A, B) - GE(A, B + 1) < 0$

Here, type a is clearly favored since the inclusion of a type a is the only option that will lead to a reduced expected error.

$$C = \frac{1}{|GE(A, B) - GE(A, B + 1)|}$$

Case 3: $GE(A, B) - GE(A + 1, B) < 0$,
 $GE(A, B) - GE(A, B + 1) > 0$

This case is symmetric to Case 2.

$$C = \frac{1}{|GE(A, B) - GE(A + 1, B)|}$$

Case 4: $GE(A, B) - GE(A + 1, B) > 0$, $GE(A, B) - GE(A, B + 1) > 0$

Here, the introduction of either type will lead to an improvement in expected group accuracy. However, since the market organizer is not soliciting agent types,

they must create a disincentive such that only one type can rationally bid. That way they can be sure of the population sizes A, B . Ideally, the agent type with the greatest contribution to the group accuracy would be the type that is not totally disincentivized from participating. However, since the C value must necessarily be negative, the type that would reduce the expected error of the group the most will always have a stronger disincentive than the opposing type. Since we would like at least one type to participate, the value for C is given by the following:

$$C = -\min\left\{\frac{1}{|GE(A, B) - GE(A + 1, B)|}, \frac{1}{|GE(A, B) - GE(A, B + 1)|}\right\}$$

i.1 Simulation for Strict Incentive Scheme

The overarching goal of the simulation is to generate a series of agents, whose error variance and covariance values satisfy type coherence, and have them choose whether or not to participate based on the incentives currently put forth by our incentive scheme. Please refer to Figure 1 for a general representation of the simulation workflow. Note that the yellow blocks indicate "setup" procedures, green blocks represent the simulation loop, and the red block indicates a halting state for the simulation.

Initialization

The initialization parameters include: $A, B, \text{var}(\epsilon_a), \text{var}(\epsilon_b), \text{cov}(\epsilon_a), \text{cov}(\epsilon_b), \text{cov}(\epsilon_a, \epsilon_b), V$. The variance and covariance values are chosen to ensure type coherence. By default, V is set to 0. These values are then used to build an $M \times M$ covariance matrix. This matrix is then squared to ensure positive semi-definiteness.

Generation of Predictions

Using the covariance matrix and the value V , the simulation constructs a multivariate normal distribution and draws from it. The draw, which is a vector of length M , represents the agents beliefs of the true value V . Note that these are not probability values and must be transformed to a vector of predictions where each entry is between 0 and 1. This transformation is done by standardizing each entry in the vector by the agent's standard deviation and mean and finding the cumulative distribution function at that value.

Market Simulation

Note that from the result by LP in [3], the expected

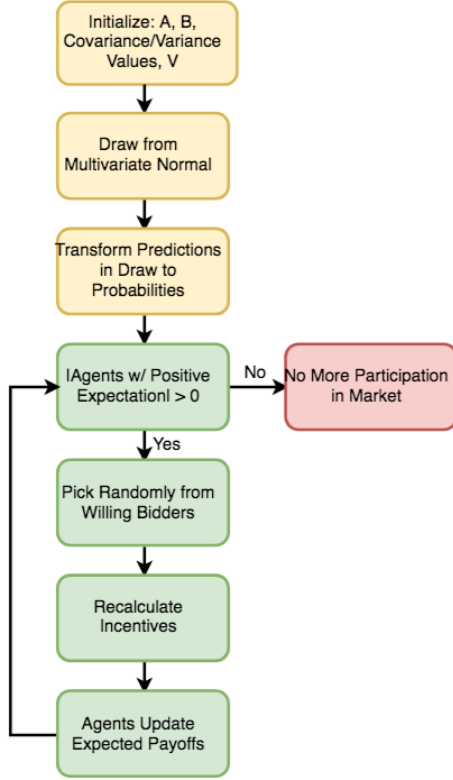


Figure 1: Simulation Workflow for Strict Incentive Scheme

group prediction error is undefined for a group of size 0 ($M = 0$). Therefore, we cannot apply our additional incentive to the first market participant. We assume arbitrarily to know the type of the first market participant. In reality, this would simply introduce an "off-by-one" error in our estimation of the expected group error.

The first participant is selected randomly from the pool of available participants, all of whom have no additional incentive or disincentive to participate. From there, the market organizer calculates the incentives and appropriate C such that the less desirable type for the current state of the market is totally disincentivized from joining per the proposed scheme. Given the newly updated set of incentives, the remaining individuals that have not yet bid recalculate their expected payoffs and choose to remain in the available bidding pool if their payoff is greater than 0. The market simulation terminates when the available pool of willing participants has size 0.

The following are a couple assumptions to note about the simulation. First, the choice to choose randomly from

the pool enforces arrival to market happens in a random order; there are no high-speed traders in this market. Second, information is integrated into the market immediately and each bidder can update their choice to participate instantaneously.

i.2 Results for Strict Incentive Scheme

Below in Figure 2, one can see participation proportions and incentives for a single simulation.

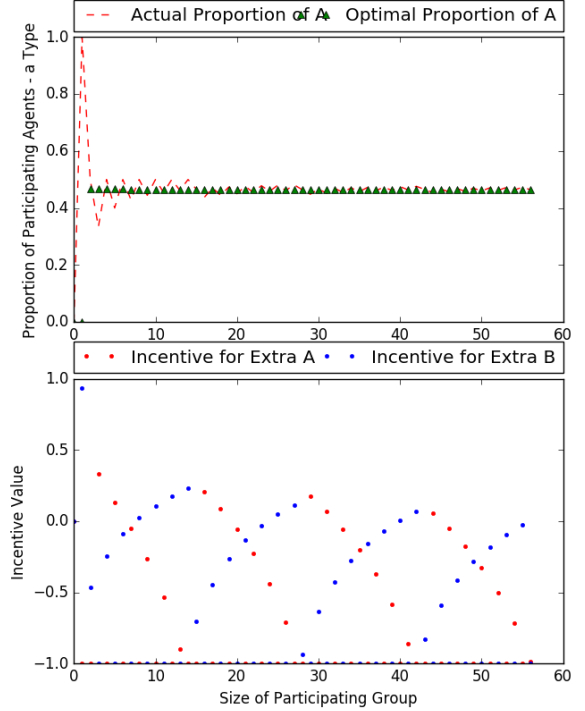


Figure 2: Market Simulation: $A=200$, $B=100$

From the simulation, one can see the incentive working in the first subplot. As the crowd gets larger, the actual proportions of the crowd adhere to the optimal proportions related to the covariance and variance statistics of the two types. However, one also notices that the market simulation terminated prior to everyone participating, since the total number of participants does not reach 300. Below, for the same point along the horizontal axis, we see that the incentives for both types for that crowd size in that simulation are both close to -1. Since the actual crowd proportions approach the optimal crowd proportions as the crowd gets larger, a slight deviation in the form of including another individual of either type will leave the crowd in a worse state. However, we know from LP [3] that

as the size of the crowd gets larger, expected group error decreases. Here, the market is halting at a local minimum expected error. To prevent this from happening, we will try to relax the incentive scheme to increase market liquidity.

ii. Relaxed Incentive Scheme

The two most salient issues with the strict incentive scheme above is the reliance on severe incentives to exclude a particular type and a lack of liquidity as the size of the participating pool increases. We would like to propose a way to relax these incentives while knowing the types of the individuals that are included in the participating pool. If the incentives for both types are greater than -1, that leaves the possibility that both types could participate. Therefore, the market organizer will need to elicit the types from the agents themselves, and the incentive scheme must guarantee that it is always in the best interests of each agent to report their type truthfully.

Previously, the scheme made the payoff for any market participant proportional to the accuracy and sharpness of their prediction as well as their effect on the expected accuracy of the group prediction. Had the individuals been asked to self-report types, there would be no incentive to either lie or tell the truth, since the incentives are not linked to their reported type but to their actual type. To ensure that there is a strict incentive to tell the truth, we propose the following addition to the scheme from earlier.

$$\pi_i(\vec{p}, \theta) = G(\vec{p}) + C[GE(A, B) - GE(A + 1, B)] - \sum_{j>i} e^{-j} \Delta GE_j \text{ if } \theta = a$$

$$\pi_i(\vec{p}, \theta) = G(\vec{p}) + C[GE(A, B) - GE(A, B + 1)] - \sum_{j>i} e^{-j} \Delta GE_j \text{ if } \theta = b$$

where j indicates an agent that participates after the current agent i participates. Furthermore, ΔGE_j is short for the change in expected group error following the inclusion of j . If $\Delta GE_j > 0$, expected group error increased, hence the subtraction of this term from the expected payoff.

This term will reward current participants in the market if future participants bring the expected group error down as well. This term provides agent i with a positive incentive to report their type truthfully. If the agent reports untruthfully, it does not affect that agent in the near term but it will skew the known proportions in the crowd such that future agents cannot make fully-informed decisions as to whether or not they are of the right type for the market

at that moment. This could lead to the inclusion of the wrong type at a particular stage in the market after i has participated.

Now that the framework is capable of reliably eliciting the truthful types from the individual agents, we are able to relax the strict bounds on the value C to allow more individuals to participate and avoid a state where no one participates. To do this, in the simulation, the market organizer will initialize C based on the conventions for the strict incentive scheme. If no agents have an expected positive payoff, decrease the magnitude of C . For example, in the simulation algorithm, C was rescaled as so:

$$C = \delta * C \quad (15)$$

where $\delta \in \mathbb{R}$ and $0 < \delta < 1$. This need not be the only correct way to scale C ; as long as $|C'| < |C|$.

ii.1 Results for Relaxed Incentive Scheme

The introduction of an adjustable C term and incentive for truthful reporting of type led to an increase in liquidity as can be seen below:

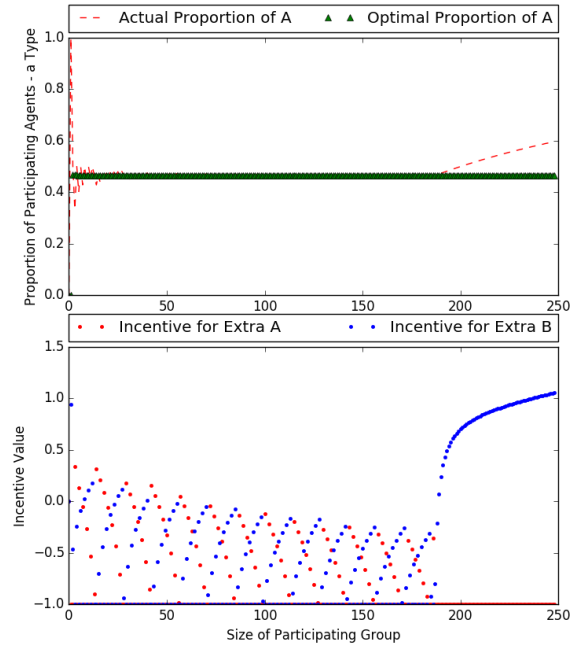


Figure 3: Market Simulation with Relaxed C : $A=200$, $B=100$

Here, the incentive scheme adapts to ensure that the market doesn't simply stop when the current proportions for the participating pool are very close or equal to the optimal proportions. One can see from the second subplot

that the incentives for type a and type b oscillate as more individuals are incorporated into the pool.

Another interesting thing to note about the two subplots is how they change when the participating group size gets to be around 180. We see a large spike in the incentive for a type b to participate as well as the proportion of the group that is of type a to increase with respect to the optimal proportion. At this point, the available type b agents have all participated.

The value of this scheme is that at any point during market operation, the market is optimally incentivizing agents that will improve the group error in expectation. With respect to accuracy, below is a graph illustrating the average prediction error of the relaxed incentive scheme versus the average error of a market without any incentive scheme.

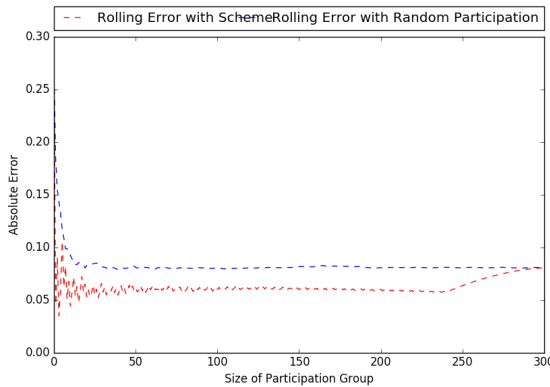


Figure 4: Accuracy Comparison for Relaxed Scheme: $A=200$, $B=100$ (average of 100 trials)

We can see that the incentive scheme helps to keep the true error low throughout the progression of bidding in the market. Notice as well that the average error for the incentive scheme begins to rise at a crowd size of 240 individuals. This sudden shift corresponds to the incentive scheme running out of bidders of the preferred type and allowing the other type to bid as well. How to detect when the market has exhausted a particular type in the available pool of participants will be discussed as a topic of future work.

ii.2 Comparison to a Market Scoring Rule

Originally, one goal was to compare the scheme proposed in this paper with the well-known market scoring rule [5] in order to compare the accuracy of one to another in determining the aggregate prediction for the likelihood of an event happening. However, there is a structural difference between the two setups that makes making a

fair comparison difficult. The results from LP in [3] rely heavily on the idea that the agents making forecasts are totally independent. However, in a market scoring rule, the utility an agent receives is linked to the last individual's prediction and oftentimes, the prediction made by an agent is altered by their reasoning over the last agent's bid. For the sake of experimentation, a market was simulated both using a market scoring rule and the proposed scheme, but the simulation revealed no significant insights.

IV. FUTURE WORK

One of the more interesting results from this paper is that the relaxed incentive scheme could maintain an optimally wise crowd at each moment of the operation of a market, up until the point where a particular type of forecaster was scarce. Given a prior distribution over the relative quantities of individual types in a prediction market, one could devise a method for shutting down a prediction market when the probability that the more desirable type has not shown up in several iterations becomes too low, indicating that the market has exhausted that perspective.

Two other topics for future work that this paper naturally leads the reader to are how the proposed incentive system changes when different (non quadratic) scoring rules are used and how much it would cost to run the incentive scheme. While it was hoped these questions could be answered in this paper, tackling the issues of thin bidding pools in the strict scheme and eliciting honest type reporting were treated with higher priority.

REFERENCES

- [1] J. Wolfers and E. Zitzewitz. Prediction markets. *The Journal of Economic Perspectives*, 18(2):107-126, 2004.
- [2] Joyce Berg, Robert Forsythe, Forrest Nelson, and Thomas Rietz. *Results from a Dozen Years of Election Futures Markets Research*, volume 1 of *Handbook of Experimental Economics Results*, chapter 80, pages 742-751. Elsevier, 2008.
- [3] P. J. Lamberson, Scott E. Page, (2012) *Optimal Forecasting Groups*. *Management Science* 58(4):805-810
- [4] Tilmann Gneiting and Adrian E. Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359-378, March 2007.
- [5] Hanson, Robin. *Combinatorial information market design*. *Information Systems Frontiers* 5.1 (2003): 107-119.