

# Forecasting with Colonel Blotto

Citation for published version (APA):

Peeters, R. J. A. P., & Wolk, K. L. (2015). *Forecasting with Colonel Blotto*. Maastricht University, Graduate School of Business and Economics. GSBE Research Memoranda No. 025  
<https://doi.org/10.26481/umagsb.2015025>

**Document status and date:**

Published: 01/01/2015

**DOI:**

[10.26481/umagsb.2015025](https://doi.org/10.26481/umagsb.2015025)

**Document Version:**

Publisher's PDF, also known as Version of record

**Please check the document version of this publication:**

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

**General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

[www.umlib.nl/taverne-license](http://www.umlib.nl/taverne-license)

**Take down policy**

If you believe that this document breaches copyright please contact us at:

[repository@maastrichtuniversity.nl](mailto:repository@maastrichtuniversity.nl)

providing details and we will investigate your claim.

Ronald Peeters, Leonard Wolk

**Forecasting with Colonel  
Blotto**

RM/15/025

**GSBE**

Maastricht University School of Business and Economics  
Graduate School of Business and Economics

P.O. Box 616  
NL- 6200 MD Maastricht  
The Netherlands

# Forecasting with Colonel Blotto\*

Ronald Peeters<sup>†</sup>

Leonard Wolck<sup>‡</sup>

September 13, 2015

## Abstract

In this paper we design and test a competitive forecasting mechanism based on the Colonel Blotto game. In the game, forecasters allocate a fixed number of resources to different ‘battlefields’. Each field is realized with a probability that is determined by a stochastic process. Subjects learn about the underlying process during the course of the experiment and thereby form beliefs about the probability that a field is selected. Once a field is selected, the subject competes for a payoff that is associated with the number of resources allocated to that field. We implement two different payment rules, a lottery and an auction, and find that the lottery outperforms the auction. This relative underperformance of the auction can be attributed to the strategic uncertainty being too high in the auction and the strong incentives to misalign allocations.

*Keywords:* forecasting; Colonel Blotto; experiment; behavioral mechanism design.

*JEL classification codes:* C72; C92; D82; D83.

---

\*We thank Matt Embrey, Georgia Kosmopoulou, Carlos Lamarche, Josh Miller, Paulo Somaini, Martin Strobel, Alexander Westkamp, and the audiences at the CEREC Workshop in Economics (Brussels, 2015), the Conference on Auctions, competition, regulation, and public policy (Lancaster, 2015) and the Behavioral and Experimental Economics Symposium (Maastricht, 2015) for very helpful comments and suggestions.

<sup>†</sup>Department of Economics, Maastricht University. [r.peeters@maastrichtuniversity.nl](mailto:r.peeters@maastrichtuniversity.nl)

<sup>‡</sup>Department of Economics, Colby College. [leonard.wolk@colby.edu](mailto:leonard.wolk@colby.edu)

# 1 Introduction

In this paper we propose a forecasting mechanism based on the Colonel Blotto game. The aim of our mechanism is to elicit forecasts of uncertain events. While belief elicitation and forecasting has been the subject of a wide range of studies to date our approach allows a forecaster to generate forecasts not only about the expected realization but also the distribution around it. We achieve this by setting up a strategic environment in which forecasters are competing for a prize.

Similar to the standard Blotto game, individuals divide resources across a set of bins, or battlefields, each of which has a different realization probability. In the game only one bin is selected and payoffs are determined based on the resources allocated to this bin, while resources allocated to other bins are lost. We show theoretically that, in a lottery payment mechanism, it is an equilibrium strategy for the players to reveal the true realization probabilities by their allocations in the game. However, when allocating payments through an auction, such an equilibrium does not generally exist.

Further, we design an experiment to test the forecasting performance of our mechanism and we show that distributions forecasted under the lottery payment outperform those from the auction setting. We next show that individuals in the auction change their allocations and coordinate their strategies to a larger extent than in the lottery. Taken together with the lower performance in forecasting, this shows that the auction mechanism results in worse forecasts since individuals allocate most of their resources to a few bins. An essential feature of our experiment is that the forecasting task is abstract and subjects have no prior exposure or knowledge that allows them to excel at the task in the first round. As a consequence we observe how subjects familiarize themselves with the environment over time and update their beliefs about the distribution of termination times.

Additionally, we elicit individual beliefs of the participants about the underlying distribution. Beliefs do not differ systematically between the auction and the lottery which shows that the beliefs held by subjects do not depend on the payment rule but that they learn about the underlying process independently of the strategies they play in the game. Thus, differences in behavior across the payment rules appear to be driven by strategic rather than environmental concerns.

Taken together, we show that it is sufficient with two competing forecasters to generate good forecasts. In many instances where relevant information is held by a select group of individuals our mechanism can help a decision maker to elicit forecasts of an uncertain event.

The mechanism presented here relates to several streams of literature. In particular, prediction markets have been developed to forecast uncertain events and have been successfully implemented to predict the outcome of political elections (Forsythe et al., 1992) and organizational processes (Cowgill & Zitzewitz, 2013) among many other applications. Yet, operating

a prediction market does not come without problems. It has been suggested that they are susceptible to manipulation (Veiga & Vorsatz, 2006) and in some countries it is difficult to operate them from a legal perspective (Arrow et al., 2008). Further, prediction markets require substantial trading to create liquid and informative market prices. Our paper presents an incentive compatible forecasting mechanism by which we are able to generate informative forecasts with fewer individuals and thus presents a viable alternative to prediction markets in environments where such markets cannot easily be operated.

An alternative to our approach that does not require a large set of traders is to elicit beliefs using scoring rules, so called ‘truth serums’ (see for instance Prelec, 2004; Trautmann & van de Kuilen, 2015). Scoring rules provide an incentive compatible mechanism for belief elicitation at an individual level and has several advantages over market based mechanisms since they are non-competitive and forecasts thereby do not suffer from strategic elements that can distort forecasts. Several studies have shown that mechanisms based on, or inspired by, scoring rules are able to generate good forecasts (see for instance Goel et al., 2010). However, one common problem is to successfully elicit distributions (Peeters & Wolk, 2015). An exception is Harrison et al. (2013) where they develop a mechanism to elicit belief distributions using an individual scoring rule. The mechanism presented in this paper explicitly deviates from the scoring rule approach by its competitive design. We explicitly introduce competition to motivate participants to exert more cognitive effort in the forecasting task. It is commonly known that games, or other competitive environments, carry a non-pecuniary value to participants, making our mechanism attractive to field adoption.

A third related stream of literature, that also incorporates the competitive element present in this paper, is the literature on parimutuel wagering, often in the context of betting (Figlewski, 1979; Thaler & Ziemba, 1988). For instance, Plott et al. (2003) design an experiment to test information aggregation using this type of mechanism. In the experiment, the authors implement a continuous market where bettors can place bets at any time prior to closing. The authors show that the mechanism induces information aggregation. Further, Koessler et al. (2012) study how beliefs interact with market odds and show that when eliciting beliefs, market odds are better aligned with the objective probabilities. Gillen et al. (2013) implement a parimutuel-like information aggregation mechanism at Intel and shows that this mechanism has predictive power against the distribution of realized outcomes. A common feature of betting markets is that they allow sequential betting. In our study all bets are placed simultaneously and players are symmetric to the extent that they possess an equal amount of resources and all have to be invested in the game in each period. The expected payment does not depend on endogenous participation in the game. Therefore, the entry decision does not depend on how many players already have entered. Additionally, we investigate two different payment rules one of which differs from the betting literature.

The paper is structured as follows. In the next sections we show the equilibrium strategy of the game under different payment rules. In Section 3 we describe the experimental setup and implementation of the game and the hypotheses of our study. In Section 4 we evaluate the performance and functioning of the mechanism using experimental data. Subsequently, in Section 5, we conclude.

## 2 Colonel Blotto meets Tullock

Let there be  $m$  events  $E_1, \dots, E_m$  with consecutive realization probabilities  $p_1, \dots, p_m$  with  $p_j > 0$  for all  $j = 1, \dots, m$  and  $\sum_{j=1}^m p_j = 1$  and we assume that these probabilities are known to the  $n$  players. Each of the  $n$  players has a unit amount of resources that are to be distributed over the  $m$  events. Let the distribution of player  $i$  be denoted by  $\sigma^i = (\sigma_1^i, \dots, \sigma_m^i)$  with  $\sigma_j^i \geq 0$  for all  $j = 1, \dots, m$  and  $\sum_{j=1}^m \sigma_j^i = 1$ . If event  $E_\ell$  realizes (which is the case with probability  $p_\ell$ ), then player  $k$  receives a share of  $\frac{(\sigma_\ell^k)^\rho}{\sum_{i=1}^n (\sigma_\ell^i)^\rho}$  of the total reward, where  $\rho > 1$  is the Tullock contest success parameter.<sup>1</sup>

For  $\rho = 1$  the share of the reward a player receives equals the probability to win in a lottery where the mass that the players put on the realized event reflects the number of tickets with which these players participate in this lottery. For  $\rho = \infty$ , the reward is evenly shared among the players that put most resources on the realized event, and the mechanism works like an auction.

**Theorem 1.** *If  $\rho < \frac{n}{n-2}$ , then, in the unique symmetric Nash equilibrium, all players distribute their resources in proportion to the realization probabilities (i.e.  $\sigma_j^i = p_j$  for all  $j = 1, \dots, m$  and for all  $i = 1, \dots, n$ ).*

**Proof.** In a Nash equilibrium  $\bar{\sigma}$ , player  $i$ 's resource distribution  $\bar{\sigma}^i$  solves

$$\max_{(\sigma_j^i)_{j=1, \dots, m}} \sum_{j=1}^m p_j \frac{(\sigma_j^i)^\rho}{(\sigma_j^i)^\rho + \sum_{\iota \neq i} (\bar{\sigma}_j^\iota)^\rho}$$

subject to  $\sigma_j^i \geq 0$  ( $j = 1, \dots, m$ ) and  $\sum_{j=1}^m \sigma_j^i = 1$

Ignoring the inequalities, the first order conditions are given by

$$p_j \frac{\rho (\bar{\sigma}_j^i)^{\rho-1} \sum_{\iota \neq i} (\bar{\sigma}_j^\iota)^\rho}{[(\bar{\sigma}_j^i)^\rho + \sum_{\iota \neq i} (\bar{\sigma}_j^\iota)^\rho]^2} = \lambda^i \quad (j = 1, \dots, m).$$

By symmetry we get, for all  $i$ ,<sup>2</sup>

$$p_j \frac{\rho (n-1)}{n^2 \bar{\sigma}_j^i} = \lambda^i \quad (j = 1, \dots, m) \quad \text{or} \quad \bar{\sigma}_j^i = \frac{1}{\lambda^i} \rho \frac{n-1}{n^2} p_j \quad (j = 1, \dots, m).$$

<sup>1</sup>It is convenient to assume that all players receive  $\frac{1}{n}$  if none of the players has put any mass on the realized event.

<sup>2</sup>Here we restrict attention to symmetric equilibria. Actually, for  $\rho = 1$  and for  $n = 2$  symmetry can be derived.

As  $\frac{1}{\lambda^i} \rho \frac{n-1}{n^2}$  is constant with respect to  $j$ , we obtain, for all  $i = 1, \dots, n$ ,

$$\bar{\sigma}_j^i = p_j \quad (j = 1, \dots, m).$$

It is easily shown that  $\rho < \frac{n}{n-2}$  is a necessary and sufficient condition for the second order conditions to be satisfied at this solution to the first order conditions.<sup>3</sup> ■

One nice feature implied by the theorem is that, in case the game designer does not know the realization probabilities, but knows that they are common knowledge to the players, equilibrium behavior would perfectly reveal these probabilities to her. For this property to be satisfied the contest parameter should not be too large relative to the number of players. For the lottery payment ( $\rho = 1$ ) we find that this feature holds for any number of players. When there are only two players, it is obtained for all finite values of the contest parameter.

On the other hand, for the auction rule ( $\rho = \infty$ ) the profile in the theorem is definitely not an equilibrium. For the auction payment rule, Roberson (2006) characterizes the unique equilibrium payoffs for the setting with two players,  $n$  battlefields, and where a player's payoff is the proportion of wins on the individual battlefields (where the winner within each battlefield is decided as in an auction). In payoff terms this situation is equivalent to our situation if our bins were all to have equal chance of being decisive. These chances not being equal in our setting further complicates a possible characterization of an equilibrium as equilibrium properties will depend highly on the realization probabilities. For example, think of a situation with two players and two bins. If the one bin has a realization probability between  $1/3$  and  $2/3$  there is an equilibrium (but it is certainly not unique) where the players allocate all their resources to different bins, while there is an equilibrium where both players put all their resources on the high probability bin in all other cases. In general, instances where a pure strategy equilibrium does not exist are easily constructed for the situation where the number of events exceeds the number of players.

### 3 Experimental design

In this section we subsequently define the decision framework, the experimental design and procedures, and the hypotheses – some of which are theory based, others are following common sense.

---

<sup>3</sup>For the two player situation with lottery payment the strategy profile in the theorem has already been shown to be the unique Nash equilibrium in Friedman (1958). As Friedman (1958) considers unequal amounts of available resources for the two players, it is pretty straight-forward to see that the  $n$ -player extension of this equilibrium strategy profile constitutes a Nash equilibrium of the game situation with  $n$  players. The uniqueness claim in the theorem is, however, not easily obtained from Friedman (1958). The proof presented here follows Robson (2005).

### 3.1 Setting

Within a group of  $n$  individuals (with  $n$  taking the values 2, 3, and 4 in our experiment), each individual is asked to allocate 100 resources over eleven bins as they are illustrated in Figure 1. These eleven bins represent the possible events (labels above the bins) of which precisely one will realize. A chance mechanism (to be explained in detail later) determines the bin that is decisive for the individuals' payoffs. That is, only the resources that the individuals allocated to this particular bin is relevant for how the total group reward of  $100n$  points is shared among the individuals.

1-10	11-20	21-30	31-40	41-50	51-60	61-70	71-80	81-90	91-100	101+
1	6	10	10	9	8	7	6	5	5	33

Figure 1: The eleven future events.

The chance mechanism that we implemented is based on a random process that starts at a value of zero at time  $t = 0$  where at each unit of time the value is incremented with a real number drawn randomly according to a normal distribution with mean zero and a fixed but unknown variance. The process terminates either when the value crosses the lower boundary at  $-2.5$ , crosses the upper boundary at  $+2.5$ , or has reached time  $t = 100$  without having reached one of these boundaries. Figure 2 shows one time series generated by this process that led to a termination at the lower bound at time  $t = 63$ . The termination time uniquely determines the bin that is decisive for the payoffs. The labels above the first ten bins in Figure 1 represent ranges of termination times; the eleventh bin represents the event that it did not terminate before  $t = 100$ .

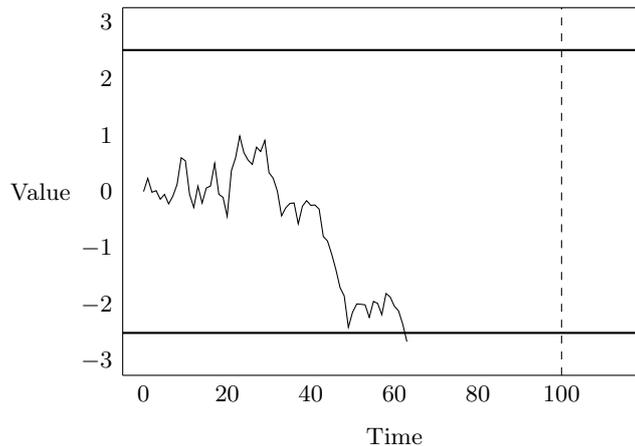


Figure 2: An example of a time series.

For the process that we used in the experiment, Figure 3 presents the true distribution over termination times, conditional on termination before  $t = 100$ . Moreover, the probability

of the process not terminating before  $t = 100$  equals one-thirds. The resources allocated in the bins of Figure 1 display the probability (rounded to the nearest integer) that the process will terminate at the respective bin.

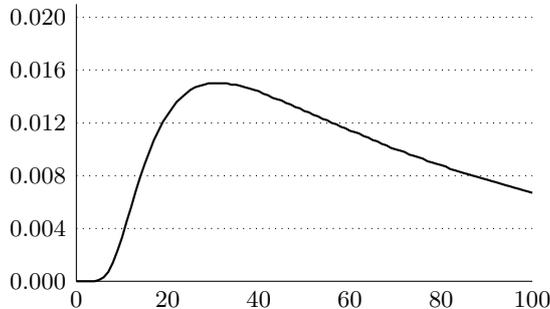


Figure 3: Distribution over termination times conditional on termination before  $t = 100$ .

We implemented two different rules on how the reward is shared in our experiment: a lottery rule and an auction rule. According to the former, the shares of the reward that the individuals receive are proportional to the amount of resources allocated to the decisive bin. In the latter, the reward is evenly distributed among the individuals that allocated the most resources to the decisive bin.<sup>4</sup>

### 3.2 Design

Individuals repeatedly interact, within the same group, over the course of twenty rounds. Prior to the first round, participants were shown an animation of a randomly generated time series. After having seen this animation, they were asked to allocate their resources over the eleven bins. Next, they were shown the animation of the time series that was generated to select the decisive bin for the first round. Prior to the second round, participants received all payoff relevant information from the first round, including the amount of resources their group members allocated to the decisive bin. This procedure continued until the last (twentieth) round.

Throughout the twenty periods participants gradually learn more about the chance mechanism determining the decisive bin. And, all participants saw the same twenty-one time series in the same order.<sup>5</sup> In the last round, before showing the final animation, we elicited

<sup>4</sup>In either case, all individuals receive an equal share of the reward when the total number of resources allocated to the decisive bin equals zero.

<sup>5</sup>In order to gain an impression of how easy it is for experimental subjects to learn about the true distribution, Figure 13 in the supplementary material displays the average Hellinger distance between the true distribution and simulated distributions on basis of a given number of realizations, and the 95% confidence interval around it. The figure reveals that the marginal reduction in the Hellinger distance is decreasing in the number of realizations and that up to about fifty realizations the reduction is substantial. Moreover, the figure shows the Hellinger distance between the true distribution and the empirical distribution on basis of the realizations throughout the experiment. The particular sequence of realizations (which was not subject to

the participants’ beliefs about the true probability distribution by which bins realize.<sup>6</sup> Final payment in the experiment was based on the points accumulated over all rounds.

Participants operated in one of the six possible treatments that varied in group size and payment rule. Table 1 summarizes these six treatments. The number of groups that were formed for each treatment was such that in total 36 participants operated in each treatment.

mechanism	group size		
	2	3	4
lottery	18 (36)	12 (36)	9 (36)
auction	18 (36)	12 (36)	9 (36)

Table 1: Number of groups (individuals) per treatment.

### 3.3 Procedures

The experiments were conducted in the experimental laboratory at Maastricht University in October 2014. We recruited undergraduate students from various disciplines via ORSEE (Greiner, 2015). All interactions took place anonymously via computer clients that were connected to a central server. The experiments were programmed in z-Tree (Fischbacher, 2007). The instructions, all screenshots, and all time series are provided in the supplementary material for one of the treatments. In total, 216 students participated in the experiment: 36 for each treatment. A typical session lasted about an hour and the average payoff was 18.26 Euros (including a 3 Euro show-up fee).

At the end of the experiment, the participants participated in a short cognition task in which we measured their perceptual reasoning ability and we elicited their risk attitude and level of Machiavellianism. For the cognition task, we used the symbol-digit correspondence test from the Wechsler Adult Intelligence Scale (WAIS), in which subjects had 90 seconds to find as many correspondences between symbols and numbers as they could, using the correct number for each symbol. Speed and accuracy under time pressure determine an individual’s ability (cf. Dohmen et al., 2010). Risk attitude was elicited by the direct approach as suggested in Dohmen et al. (2011). The Machiavellianism score was elicited using the Mach-IV test (Christie & Geis, 1970). Moreover, we elicited a few personal characteristics, including gender and age.

---

experimental manipulation) allows a better than average learning. The fact that the time series were shown as animation ensured that experimental subjects did not see the final realization only but all gradual innovations leading to a termination time – this facilitating an even better learning process.

<sup>6</sup>We incentivized them in such a way that they received one point for each probability point well assigned.

### 3.4 Hypotheses

Our main hypothesis is formulated on basis of the extent to which the distribution of resources aggregated on group level reflect the true realization probabilities. We call this the forecasting performance. According to Theorem 1, which is formulated while assuming common knowledge of these probabilities, forecasting performance is independent of the group size for the lottery payment rule. For the auction format we are not able to characterize the equilibria.<sup>7</sup> Though, as it is totally in alignment with the preferences that aggregated resources are allocated monotonically increasing in realization probability, and despite knowing that there does not exist an equilibrium that shares the elegant property of the equilibrium in the lottery, we take equal performance across treatments as working hypothesis:

**Hypothesis 1.** *Forecasting performance is independent of the payment rule and group size.*

Unlike in the lottery payment, in the auction payment resources allocated to bins to which a rival player allocated more can be considered as waste. After all, even if the respective bin is selected these resources do not yield any payoff in the auction, while they always yield something in the lottery, no matter how small the own amount allocated or how large the amount allocated by the other players. Therefore, we expect more changes in allocations over time in the auction than in the lottery on the individual level. Theorem 1 provides no reason to expect any impact of group size on changes in the lottery. One prominent reason to expect an impact of group size on changes in the auction format is that changes may be positively related to the level of strategic uncertainty – the level of strategic uncertainty being larger for larger group sizes.<sup>8</sup> This all yields the following working hypothesis:

**Hypothesis 2.** *Individual allocations change more in the auction than in the lottery. In the lottery, there is no impact of group size on the number of changes. In the auction, changes are, if anything, increasing in group size.*

Based on roughly the same arguments leading to previous hypothesis, we can formulate the next hypothesis concerning the disagreement across individual allocations within groups:

**Hypothesis 3.** *Within groups, there is more disagreement in individual allocations in the auction than in the lottery. In the lottery, there is no impact of group size on the disagreement in allocations within groups. In the auction, if anything, there will be more disagreement for larger group sizes.*

---

<sup>7</sup>Numerically solving for an equilibrium of the discretized game as implemented in the experiment is computationally hard given the high number of actions that players have at their disposal: there are  $\frac{(k+m-1)!}{(m-1)!k!}$  possible ways to allocate  $k$  resources over  $m$  bins, which amounts to almost 47 trillion actions per player in our experiment.

<sup>8</sup>To explain, when  $n = 2$ , losing on a particular bin implies a potential high probability of gaining on another bin. When  $n = 4$ , this implication is less straightforward. Stated differently, with more opponents there are more individuals to learn about, and to respond to.

In our experiment, all individuals experience the same realizations of the time series in the same order in all treatments. From that perspective there is no reason to expect that individuals will develop different beliefs in different treatments. And, even in case differences in the level of strategic uncertainty may obstruct belief formation process, there is no reason to expect individuals within the same group to develop different beliefs. This gives rise to our final hypothesis:

**Hypothesis 4.** *The level of disagreement in beliefs within groups is independent of the payment rule and group size.*

## 4 Results

For both the treatments with lottery and auction payment, we collected the sequence of twenty allocation decisions for eighteen, twelve and nine different groups consisting of respectively two, three and four members. For all statistical tests we use group averages as unit of observation. Although for aesthetical reasons the figures restrict focus to the first ten bins, all statistics are based on all eleven bins. Table 2 provides summary statistics on our pool of participants.

	all	lottery			auction		
		$n = 2$	$n = 3$	$n = 4$	$n = 2$	$n = 3$	$n = 4$
gender (% male = 1)	47.7%	36.1%	52.8%	41.7%	61.1%	63.9%	30.6%
age (years)	21.1 (2.7)	21.4 (2.2)	21.4 (2.3)	20.2 (4.0)	21.2 (2.1)	22.0 (2.9)	20.4 (2.0)
risk attitude (0–10)	6.3 (1.9)	6.2 (1.8)	6.4 (1.7)	6.2 (2.1)	6.6 (1.9)	6.3 (1.9)	6.2 (2.0)
cognitive ability (number)	42.6 (7.1)	42.5 (6.9)	41.9 (5.8)	43.2 (5.4)	41.7 (7.2)	43.4 (6.8)	43.0 (9.9)
Machiavellianism (20–100)	58.2 (7.2)	58.7 (6.5)	59.8 (7.4)	57.0 (7.4)	57.2 (6.9)	57.7 (7.3)	59.0 (7.6)

Table 2: Summary statistics (mean values and standard deviation in parentheses) of the participants in the experiment.

### 4.1 Forecasting performance

Figure 4 presents the average fraction of resources allocated to the different bins in the first round (dashed) and last round (solid) for the treatments where  $n = 3$ , taking into account the first ten bins only. The left graph shows these for the lottery; the right one for the auction. The gray-shaded area presents the true distribution (conditional on the first ten bins). From the graph it is visible that over rounds the resource allocations moved in the direction of the

true distribution.<sup>9</sup>

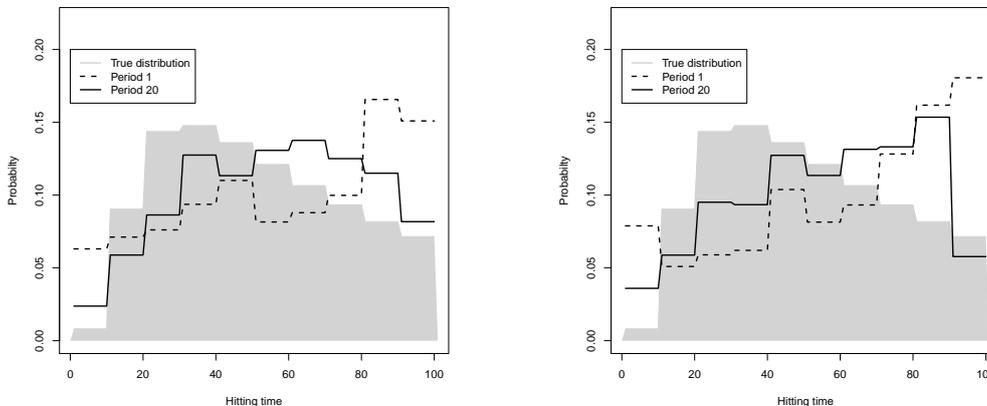


Figure 4: Forecasted distributions (conditioned on first ten bins).

In order to formally assess the impact of payment rule and group size on the performance of the group allocations as forecasts it is important to adopt a good measure. Like in Peeters and Wolk (2015), we use the Hellinger distance (Hellinger, 1909) between the group allocation ( $\tilde{\sigma}$ ) and the true distribution ( $p$ ):

$$H(\tilde{\sigma}, p) = \frac{1}{\sqrt{2}} \sqrt{\sum_{j=1}^m (\sqrt{\tilde{\sigma}_j} - \sqrt{p_j})^2}.$$

This measure quantifies the similarity between two probability distributions and takes the maximal value of one in case the supports of the group allocation and the true distribution are disjoint and the minimal value of zero when the two distributions are identical.<sup>10</sup>

Figure 5 shows the performance of the two mechanisms for the various group sizes. In all treatments we see a steep decline in the first three rounds, identifying that the way in which resources are allocated over bins on group-level become more in line with the true probability distribution. From the fourth round onwards, we see a further declining trend in the lottery while in the auction there is no further structural decline for any group size. All differences across treatments are summarized by the following result.

**Result 1.** *The lottery outperforms the auction for all group sizes. For the lottery, performance significantly increases from  $n = 2$  to  $n = 3$  and decreases (not significantly though) from  $n = 3$  to  $n = 4$ . For the auction, there is no impact of group size on performance.*

<sup>9</sup>A similar movement is found in the graphs for the treatments with  $n = 2$  and  $n = 4$ ; see Figure 11 in the supplementary material.

<sup>10</sup>An important advantage of the Hellinger distance over often used alternatives (such as the Kullbeck-Leibler divergence) is that it does not require absolute continuity. Another desirable property of the Hellinger distance, that we do not exploit here, is that it satisfies the triangular inequality. One notable shortcoming of the Hellinger distance, which is shared with all existing alternative measures, is that it does not take into account the linear order on the domain.

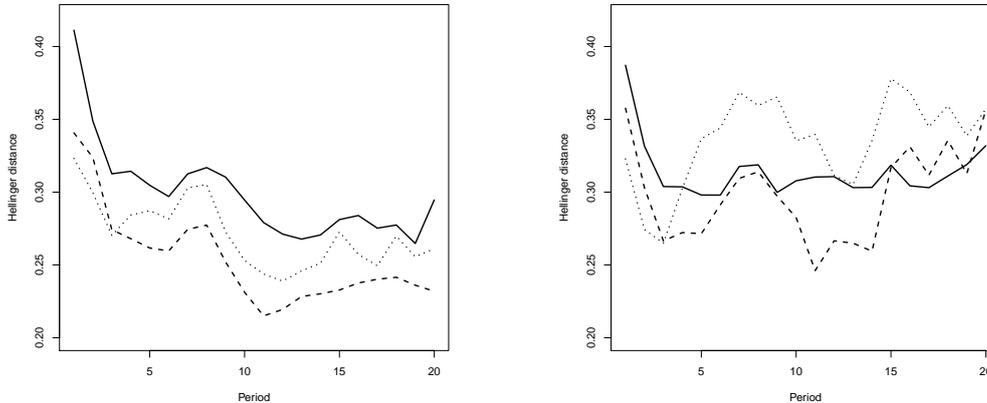


Figure 5: Average Hellinger distance between group allocation and the true distribution over time for the different treatments. Left: lottery; Right: auction. Solid:  $n = 2$ ; Dashed:  $n = 3$ ; Dotted:  $n = 4$ .

**Support.** We test for treatment differences using two-sided Mann-Whitney U tests taking for each group the average Hellinger distance between the group allocation and true distribution over all rounds as unit of observation. The  $p$ -values of the tests comparing the lottery and the auction are .481, .010 and .000 for  $n = 2$ ,  $n = 3$  and  $n = 4$ , respectively. For the lottery, the  $p$ -value of the test comparing  $n = 2$  with  $n = 3$  equals .001, the one comparing  $n = 3$  with  $n = 4$  equals .219, and the one comparing  $n = 2$  with  $n = 4$  equals .020. For the auction, the  $p$ -values of the respective comparisons equals .917, .219, and .194. At the 95% confidence level, the conclusions do not change when basing the analysis on observations in the last ten rounds; the only thing worth noticing is that the  $p$ -value when comparing lottery and auction for  $n = 2$  drops to .055. ■

So, compared to the auction, the lottery is the preferred mechanism. In addition, as can be seen in the graph, there is less volatility in performance in the lottery – making the lottery a more robust mechanism.<sup>11</sup> One explanation for the observed volatility in the auction is that in addition to the environmental uncertainty (about realization probabilities) players face a high level of strategic uncertainty that is caused by the competitiveness of the mechanism providing incentives to misalign allocations. This we further explore in the subsections to follow. Furthermore, while within the lottery increasing the number of participants initially increases performance, beyond a certain number of participants we find no additional benefit of adding further participants. In our experiment this threshold was found at three players.

<sup>11</sup>Also the variance in forecasting performance across groups is smaller in the lottery than in the auction – this effect being significant for group sizes of three and four (see Section 6 of the supplementary material). This shows the lottery being more robust as a forecasting mechanism compared to the auction.

## 4.2 Changes in decisions

In order to quantify the level of change in an individual's allocation of resources from one round ( $\sigma_{t-1}^i$ ) to the other ( $\sigma_t^i$ ), we use the Hellinger-distance between these two allocations ( $H(\sigma_t^i, \sigma_{t-1}^i)$ ). Figure 6 shows the intensity of changes over rounds for the two mechanisms and the various group sizes. In all treatments we see a steep decline in the first three to four rounds. From there onwards, we see a further declining trend in the lottery while in the auction there is no further structural decline for any group size. So, while in the auction they tend to redistribute a fixed fraction of the total resources available from round to round, in the lottery they redistribute less resources over rounds.

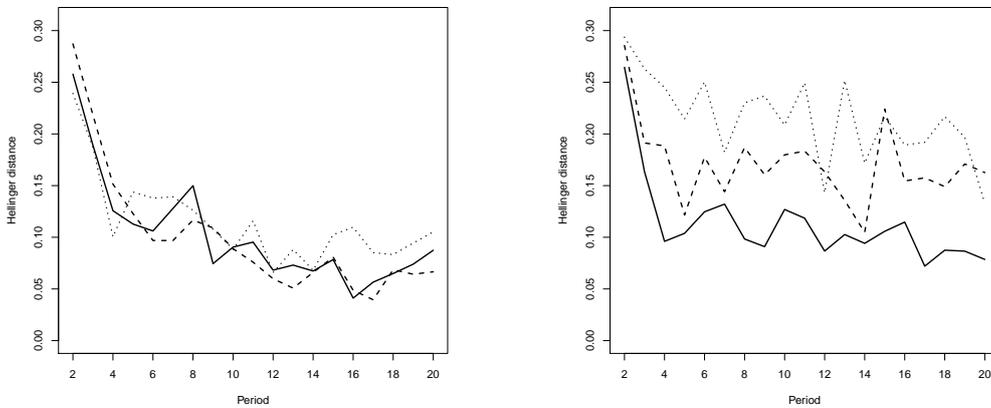


Figure 6: Changes in individual allocation of resources. Left: lottery; Right: auction. Solid:  $n = 2$ ; Dashed:  $n = 3$ ; Dotted:  $n = 4$ .

**Result 2.** *Individuals change allocations more in the auction for groups of size three and four; there is no significant difference between lottery and auction for groups of size two. For the lottery, there is no significant impact of group size on allocation changes on individual level. For the auction, there are more changes in  $n = 3$  and  $n = 4$  compared to  $n = 2$  and no significant difference between  $n = 3$  and  $n = 4$ .*

**Support.** We test for treatment differences using two-sided Mann-Whitney U tests taking for each group the average Hellinger distance between the individual allocation in round  $k$  with that in  $k - 1$  over all rounds as unit of observation. The  $p$ -values of the tests comparing the lottery and the auction are .542, .005 and .006 for  $n = 2$ ,  $n = 3$  and  $n = 4$ , respectively. For the lottery, the  $p$ -value of the test comparing  $n = 2$  with  $n = 3$  equals .491, the one comparing  $n = 3$  with  $n = 4$  equals .345, and the one comparing  $n = 2$  with  $n = 4$  equals .561. For the auction, the  $p$ -values of the respective comparisons equals .012, .219, and .001. At the 95% confidence level, the conclusions do not change when basing the analysis on

observations in the last ten rounds. ■

A significant difference in individual changes between two treatments may not imply that also changes in allocations on group level will be different. For example, two individuals may make precisely the opposite marginal change – this implying that two individuals change, while nothing changes on group level. Figure 7 shows the average Hellinger distance between the group allocations in two consecutive rounds ( $H(\tilde{\sigma}_t, \tilde{\sigma}_{t-1})$ ). Not only do the graphs look quite similar to those for individual changes, also the same conclusions with respect to treatment differences can be drawn (Result 2).<sup>12</sup>

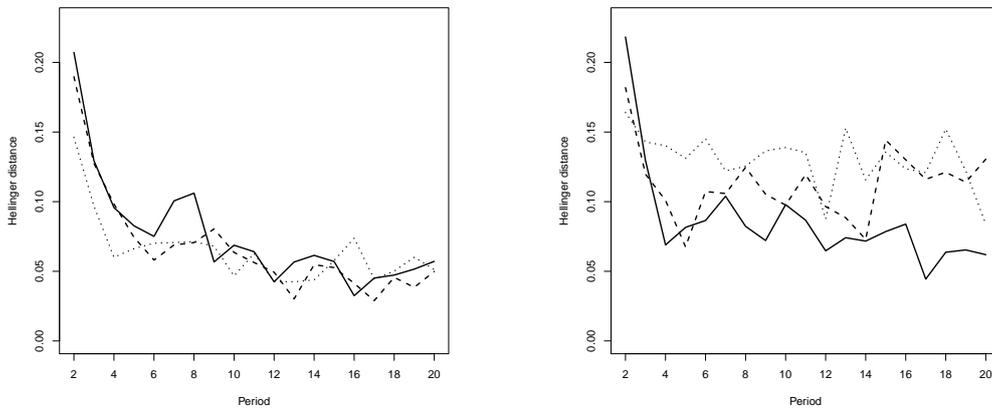


Figure 7: Changes in individual allocation of resources. Left: lottery; Right: auction. Solid:  $n = 2$ ; Dashed:  $n = 3$ ; Dotted:  $n = 4$ .

Based on these findings we may conclude that the volatility in the performance of the auction (relative to the lottery) is caused by individuals changing their allocations a lot. Yet, it is not to be concluded that it drives the significant difference in the level of performance across mechanisms.

### 4.3 Within group agreement

One explanation for there being more changes in the auction compared to the lottery is that in the auction players have an incentives to concentrate their resources on a partition of the bins that is ignored by the other players. This implies that the individual allocations within a group are disagree more in the auction compared to the lottery. We suggest the following index to measure the level of disagreement within groups.

$$DA((\sigma^i)_{i=1}^n) = \frac{1}{n} \sum_{i=1}^n H(\tilde{\sigma}, \sigma^i).$$

<sup>12</sup>The  $p$ -values of the nine tests in the evidence are respectively .563, .010, .004, .185, .651, .433, .072, .422 and .059. Moreover, the two  $p$ -values that are just above .05 drop to below .05 (.048 and .031 respectively) when restricting attention to the decisions in the last half of the experiment.

This agreement index entails the average Hellinger distance between the individuals' allocations and their group allocation and satisfies several desirable axioms including continuity, anonymity, normalization and replication invariance (which allows a fair comparison of the index when applied to groups with different numbers of participants).<sup>13</sup>

Figure 8 displays the development of the disagreement in allocations within groups over rounds for the six treatments. Like in earlier figures the left-hand panel shows these for the lottery and the right one for the auction; the solid, dashed and dotted lines refer to averages for groups of size two, three and four, respectively. While for the lottery we see again an inclining trend in the first few round, we do not see this pattern for the auction. Moreover, after an initial phase, the disagreement in the lottery seems to be declining while it is increasing in the auction. The earlier mentioned incentives to miscoordinate in the auction, that are not as prominent in the lottery, can explain the difference in these trends.

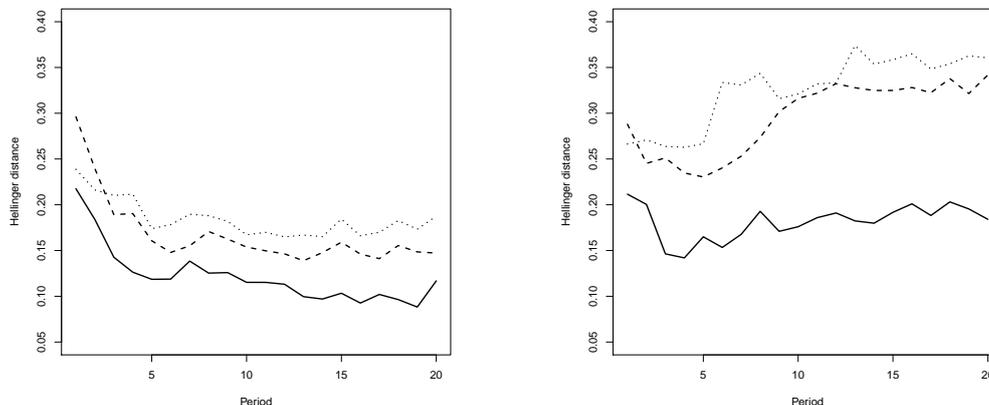


Figure 8: Disagreement in allocations within groups. Left: lottery; Right: auction. Solid:  $n = 2$ ; Dashed:  $n = 3$ ; Dotted:  $n = 4$ .

**Result 3.** *For all group sizes, there is more agreement across allocations within groups in the lottery than in the auction. For the lottery, agreement is decreasing in group size (not significant for marginal changes). For the auction, agreement is decreasing in group size as well.*

**Support.** We test for treatment differences using two-sided Mann-Whitney U tests taking for each group the average disagreement index over all rounds as unit of observation. The  $p$ -values of the tests comparing the lottery and the auction are .001, .002 and .000 for  $n = 2$ ,  $n = 3$  and  $n = 4$ , respectively. For the lottery, the  $p$ -value of the test comparing  $n = 2$  with  $n = 3$  equals .439, the one comparing  $n = 3$  with  $n = 4$  equals .095, and the one comparing

<sup>13</sup>These axioms are part of the system to characterize indices for multi-dimensional inequality measurement (Tsui, 1999) and measuring cohesiveness of preferences (Alcalde-Unzu & Vorsatz, 2013). Other axioms that are used in these respective literatures seem less suitable to adopt in the present situation.

$n = 2$  with  $n = 4$  equals .004. For the auction, the  $p$ -values of the respective comparisons equals .000, .219, and .000. At the 95% confidence level, the only change when basing the analysis on observations in the last ten rounds is in the between  $n = 3$  and  $n = 4$  in the auction of which the  $p$ -value drops to .049. ■

The individuals in our experimental population clearly realized the strong incentives to mis-align allocations in the auction, and these incentives seem to be stronger when the number of participating players increases. This increases the level of strategic uncertainty in the auction relative to the lottery. We believe it is precisely this strategic uncertainty that causes the lower forecasting performance.

#### 4.4 Environmental versus strategic uncertainty and common knowledge

In this final subsection we first try to gain understanding of the motives behind individual allocations, by considering their beliefs and contrasting these with their allocations. Next, we assess in how far individuals are able to form a common belief on the true realization probabilities.

Figure 9 presents the average fraction of resources allocated to the different bins in the last round (solid) and the average belief for the realization probabilities of each bin as revealed after the last round (dashed) for the treatments where  $n = 3$ , taking into account the first ten bins only. The left graph shows these for the lottery; the right one for the auction. The gray-shaded area presents the true distribution (conditional on the first ten bins). From these graphs it appears that allocations and beliefs are less aligned in the auction relative to the lottery.<sup>14</sup>

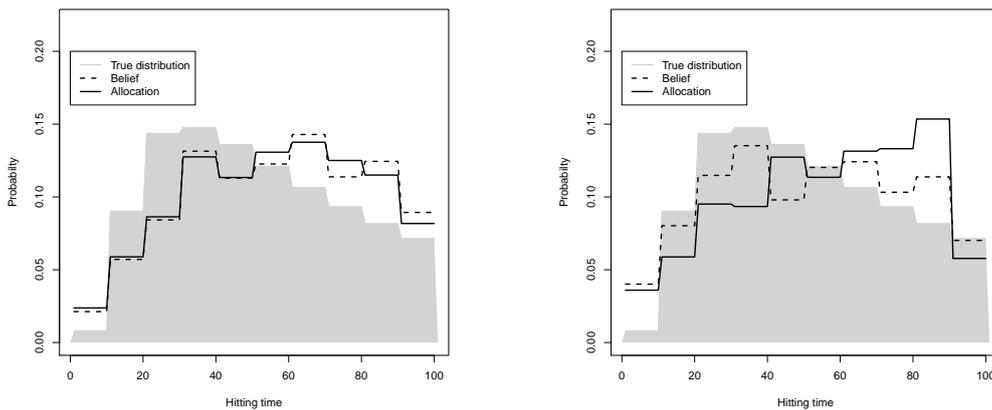


Figure 9: Average beliefs (dashed) and allocations (solid) in the last round, compared against the true distribution (grey scale); all conditioned on first ten bins. Left: lottery; Right: auction.

<sup>14</sup>A similar effect is present in the graphs for the treatments with  $n = 2$  and  $n = 4$ ; see Figure 12 in the supplementary material.

Table 3 presents for each treatment (1) the average Hellinger distance between individual beliefs and the true realization probabilities, (2) the average Hellinger distance between individual beliefs and allocations, and (3) the average disagreement across individual beliefs within groups.

mechanism	group size								
	2			3			4		
	$H(\beta^i, p)$	$H(\beta^i, \sigma^i)$	$DA(\beta)$	$H(\beta^i, p)$	$H(\beta^i, \sigma^i)$	$DA(\beta)$	$H(\beta^i, p)$	$H(\beta^i, \sigma^i)$	$DA(\beta)$
lottery	0.3248	0.2323	0.1326	0.2916	0.2012	0.1582	0.3300	0.2367	0.1944
auction	0.3903	0.3200	0.2128	0.5035	0.4049	0.2229	0.5188	0.4597	0.2231

Table 3: Quality of individual beliefs ( $H(\beta^i, p)$ ), agreement between individual beliefs and allocations ( $H(\beta^i, \sigma^i)$ ), and agreement of beliefs within groups ( $DA(\beta)$ ).

In the lottery, individuals form better beliefs in groups of size three than in groups of size two or four – only the comparison between groups of size three and four is significant (the relevant  $p$ -values are .104, .023 and .781). In the auction, individuals’ beliefs are worse the larger the group size – only the comparison between groups of sizes three and four not being significant (the relevant  $p$ -values are .000, .651 and .001). Moreover, individuals form significantly better beliefs on what the true realization probabilities are in the lottery than in the auction for all group sizes (the relevant  $p$ -values are .014, .000 and .000). From all these findings it seems that an increase in strategic uncertainty comes at the expense of learning about the uncertainty that is coming from the environment.<sup>15</sup>

In the lottery, individuals’ allocations of resources in the last round of play are closer to their beliefs in groups of size three than in groups of size two or four – though not significant in any comparison (the relevant  $p$ -values are .171, .310 and .980). Although insignificant in effect, this inverse-u-shaped relation seems sufficient to amplify the similar inverse-u-shaped relation on the belief formation, to generate the inverse-u-shaped relation in forecasting performance in Result 1. In the auction, individual allocations are less similar to individual beliefs the larger the group size – though not significant for the marginal comparisons (the relevant  $p$ -values are .072, .219 and .017). While beliefs became worse in group size, allocations seem to counterbalance this effect, in order to destroy any possible impact of group size on forecasting performance (Result 1). Finally, and as expected, individual allocations of resources in the last round of play are significantly closer to the individuals’ beliefs in the lottery than in the auction (the relevant  $p$ -values are .020, .000 and .001). We believe the main reason of this is

<sup>15</sup>In Table 5 of the supplementary material we present the Hellinger distance between individual beliefs and the distribution based on the realizations up to and including period 19, the period preceding the period that we elicited the individual beliefs. In the lottery the stated beliefs are substantially closer to the true beliefs than to the realizations. For the auction we find a similar property for groups of size two; for larger groups we find that the beliefs are negligibly closer to the realizations than to the true beliefs. This seems to exclude that beliefs are biased by the realized termination times – possibly indicating that all gradual innovations are considered informative as well by the participants.

not the difference in the belief formation process, but the incentive to misalign allocations (Result 3).

Both in the lottery and the auction, there is less agreement across beliefs within groups for larger groups sizes, but none of the marginal effects is significant (the relevant  $p$ -values are .662, .129 and .007 for the lottery and .692, .917 and .743 for the auction). Furthermore, there is more agreement across beliefs within groups in the lottery than in the auction for all group sizes – only the comparison for group size of four not being significant (the relevant  $p$ -values are .022, .024 and .863).

**Result 4.** *For all group sizes, there is more agreement across beliefs within groups in the lottery than in the auction. Both in the lottery and the auction, commonality in belief is not highly dependent on group size.*

So, the common belief assumption that what made in Section 2 finds less support in the auction relative to the lottery. There being a difference is striking since all individuals in all treatments experienced the same realized time series in all rounds. Apparently, competitiveness of the payment rule has an impact on the learning process.

## 5 Conclusion

In this paper we develop and experimentally test a strategic mechanism to generate forecasts of an uncertain event. We consider two different payment rules, an auction and a lottery. In the lottery mechanism there is a unique equilibrium to reveal ones beliefs while this is not the case under the auction payoff rule. To test the two payment rules we design an experiment where subjects are asked to forecast the termination time of a timeseries. Unique to our design is that subjects observe how this series develops dynamically over time and the subjects thereby form beliefs about the underlying process without any precise quantitative information. Our experimental results show that over time individuals in the lottery treatment are able to forecast the underlying distribution with a significantly higher precision than in the auction.

The findings presented here are in line with a growing literature on non-market based elicitation mechanisms. Indeed, our results show that even with less than five forecasters we are able to construct good distributional forecasts. Under the lottery payment rule we further observe that three forecasters appear better than both two and four. Not having to rely on many forecasters is essential when thinking about implementing forecasting mechanisms in corporate settings where information may be highly sensitive and sharing of such information unwanted.

Taken together, our results have important implications for the design of non-market based forecasting mechanisms. First, we show that the payment rule significantly affects the

equilibrium strategy of the game and also the observed behavior in the experiment. Second, we show that this is largely an artifact of the increased strategic uncertainty of the game when played with the auction rule. Under this rule, subjects are trying to outdo their opponents by constantly changing their allocations across rounds so as to have the upper hand in the game. This is evident when both looking at the individual changes across rounds and also when looking at the within group disagreement.

Consequently, setting up competitive environments to elicit forecasts where the payment rule allocates the entire payoff to the realized winner does not necessarily induce truthful revelations of beliefs and will therefore lead to lower quality forecasts. Our paper suggests that a lottery-type payment rule offers better distributional forecasts and should be preferred over tournament like payment rules.

## References

1. Alcalde-Unzu, J. and M. Vorsatz (2013). Measuring the cohesiveness of preferences: An axiomatic analysis. *Social Choice and Welfare* 41(4): 965-988.
2. Arrow, K.J., R. Forsythe, M. Gorham, R. Hahn, R. Hanson, J.O. Ledyard, et al. (2008). The promise of prediction markets. *Science* 320(5878): 877-878.
3. Christie, R. and F. Geis (1970). How devious are you? Take the Machiavelli test to find out. *Journal of Management in Engineering* 15.4: 17.
4. Cowgill, B. and E. Zitzewitz (2013). Corporate prediction markets: Evidence from Google, Ford and Firm X. Working paper.
5. Dohmen T., A. Falk, D. Huffman and U. Sunde (2010). Are risk aversion and impatience related to cognitive ability? *The American Economic Review* 100(3): 1238-1260.
6. Dohmen T., A. Falk, D. Huffman, U. Sunde, J. Schupp and G. Wagner (2011). Individual risk attitudes: Measurement, determinants and behavioral consequences. *Journal of the European Economic Association* 9(3): 522-550.
7. Figlewski, S. (1979). Subjective information and market efficiency in a betting market. *Journal of Political Economy*. 87(1): 75-88.
8. Fischbacher, U. (2007). zTree: Zurich toolbox for ready-made economic experiments. *Experimental Economics* 10(2): 171-178.
9. Forsythe, R., F. Nelson, G.R. Neumann and R. Wright (1992). Anatomy of an experimental political stock market. *American Economic Review* 82(5): 1142-1161.
10. Friedman, L. (1958). Game-theory models in the allocation of advertising expenditures. *Operations Research* 6(5): 699-709.
11. Gillen, B.J., C.R. Plott and M. Shum (2013). Inside Intel: Sales forecasting using an information aggregation mechanism. Working paper.
12. Goel, S., D.M. Reeves, D.J. Watts and D.M. Pennock (2010). Prediction without markets. *Proceedings of the ACM EC'10 Conference, Cambridge, MA*.
13. Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with ORSEE. *Journal of the Economic Science Association*, 1(1): 114-125.
14. Harrison, G.W., J. Martinez-Correa, T. Swarthout and E.R. Ulm (2013). Scoring rules for subjective probability distributions.

15. Hellinger, E. (1909). Neue Begründung der Theorie quadratischer Formen von unendlichvielen Veränderlichen. *Journal für die reine und angewandte Mathematik* 136: 210-271.
16. Koessler, F., C. Noussair and A. Ziegelmeyer (2012). Information aggregation and belief elicitation in experimental parimutuel betting markets. *Journal of Economic Behavior and Organization* 83: 195-208.
17. Peeters, R. and L. Wolk (2015). Eliciting interval beliefs: An experimental study. *GSBE Research Memorandum* 14/029.
18. Plott, C.R., J. Wit and W.C. Yang (2003). Parimutuel betting markets as information aggregation devices: experimental results. *Economic Theory* 22: 311-351.
19. Prelec, D. (2004). A Bayesian truth serum for subjective data. *Science* 306(5695): 462-466.
20. Roberson, B. (2006). The Colonel Blotto game. *Economic Theory* 29: 1-24.
21. Robson, A. (2005). Multi-item contests. Working paper (January, 2005).
22. Thaler, R. and W.T. Ziemba (1988). Parimutuel betting markets: Racetracks and lotteries. *Journal of Economic Perspectives*. 2(2): 161-174.
23. Trautmann, S.T. and G. van de Kuilen (2015). Belief Elicitation: A Horse Race among Truth Serums. *Economic Journal*.
24. Tsui, K-Y. (1999). Multidimensional inequality and multidimensional generalized entropy measures: An axiomatic derivation. *Social Choice and Welfare* 16(1): 145-157.
25. Veiga, H. and M. Vorsatz (2006). Price manipulation in an experimental asset market. *European Economic Review* 53: 327-342.