A Field Experiment on Monetary Incentives in Prediction Markets

Stefan Luckner
Institute IISM
Universität Karlsruhe (TH)
Karlsruhe, Germany
luckner@iism.uka.de

ABSTRACT
The results of recent studies on prediction markets are encouraging. Prior experience demonstrates that markets with different incentive schemes predicted uncertain future events at a remarkable accuracy. In this paper, we study the impact of different monetary incentives on the prediction accuracy in a field experiment. In order to do so, we compare three groups of users, corresponding to three treatments with different incentive schemes, in a prediction market for the FIFA World Cup 2006. Somewhat surprisingly, our results show that performance-compatible payment does not necessarily increase the prediction accuracy.

Categories and Subject Descriptors
H.3.5 [**Information Storage and Retrieval**]: Online Information Services – *commercial services, web-based services*; J.4 [**Social and Behavioral Sciences**]: *Economics*

General Terms
Design, Economics, Experimentation.

Keywords
Prediction Markets, Incentive Engineering, Field Experiment.

1. INTRODUCTION
Prediction markets are a promising approach for forecasting uncertain future events. The basic idea of a prediction market is to trade virtual stocks with certain payoffs that depend on uncertain future events. Examples comprise the outcome of an election or the results of a sports event. The Iowa Electronic Market (IEM) for predicting the outcome of the presidential elections in 1988 was the first political stock market [2]. Since then, political stock markets have been widely used as an alternative to polls and initially seemed to be the miracle cure in psephology. Apart from political stock markets, the idea behind prediction markets has also been used in various settings like in market research or business forecasting in general [12, 13]. Lately, forecasting markets are also used in order to predict the outcome of sports events [7].

The principle idea is that according to the efficient market hypothesis [1], prices of traded assets reflect all available information and, thus, asset prices can be used to predict the likelihood of uncertain events. Consider a share that promises a payment of one currency unit for every percentage point a party obtains at an election. If, for example, a party wins 40 percent at the election, the participants receive 40 currency units for each share of that party they have in their portfolio. An investor who believes that the party will obtain 40 percentage points might sell his shares of this party for prices above and buy additional shares for prices below 40 currency units. Thus, the market prices reflect the expectations of the traders regarding the outcome of the election [8]. Several studies have shown that the market prices of the shares prior to the election are very close to the percentage points the respective parties win at the actual election.

The focal point of this work is to study the impact of different incentive schemes on the prediction accuracy in a field experiment. We want to elaborate on the question whether prediction markets with performance-related incentives perform better than markets with fixed payments. Somewhat surprisingly, our results show that performance-compatible incentives do not necessarily increase the prediction accuracy. Based on our results we will give advice on engineering incentive schemes for future prediction markets.

The remainder of the paper is structured as follows: The next section describes some related work on incentives schemes in the area of experimental economics and two studies on real-money vs. play-money prediction markets. In section 3, we then describe the setup of our field experiment we conducted during the FIFA World Cup 2006 in Germany. Furthermore, we discuss our results concerning the impact of different incentive schemes on the prediction accuracy in section 4. Thereby, we also speculate why using performance-related incentives could possibly lead to a decrease in prediction accuracy. In section 5, we finally summarize our findings and give an outlook on possible implications these results might have on designing incentive schemes for public and intra-enterprise prediction markets.

2. RELATED WORK

Previous research in the field of prediction markets has shown that play-money as well as real-money markets can predict future events at a remarkable accuracy [2, 13]. So far, market operators have employed various kinds of incentive schemes in order to motivate people to take part in such markets and to reveal their expectations. Typical examples are prizes for the top performers of a market, lotteries among all traders, rankings published on the Internet or even real-money exchanges. We suspect that the embodiment of the incentive mechanisms has a huge impact on the market quality and the prediction accuracy. Despite this, we are aware of merely two papers studying incentives for prediction markets by comparing real-money and play-money markets.

In one of these two earlier studies, Servan-Schreiber et al. found that there was no statistically significant difference between the real-money market TradeSports and the play-money market NewsFutures [11]. Rosenbloom et al., however, found TradeSports to be significantly more accurate than NewsFutures for non-sports events [10]. In case of NFL games, they produced conclusions consistent with those from Servan-Schreiber et al. Considering both studies, we believe that the impact of real-money vs. play-money still remains an open question in the field of prediction markets. Moreover, there exists far more than one design option only for play-money markets – and also for real-money markets. The strength of both studies is the large data set from real-world online experiments that both papers rely on. However, both studies do not consider any other differences apart from the use of real-money or play-money in their comparison of the two markets. Although the markets they compare are quite similar, they are by far not identical. We agree that a key difference between the two markets is that one uses real-money while the other does not. But how did some other aspects influence the prediction accuracy? It remains an open question how e.g. the number of traders and their trading activity influences the market and thus also the prediction accuracy. This seems to be an interesting question, since the number of traders per contract was not available for TradeSports. What is more, TradeSports does also levy a small fee on each transaction. How does this impact the trading behavior and the resulting share prices? The two markets – TradeSports and NewsFutures – were not identical and we thus claim that other influencing factors might have caused the results described by Servan-Schreiber et al. and also by Rosenbloom et al.

As already mentioned before, these two are the only papers dealing with incentive schemes that we are aware of in the field of prediction markets. In experimental economics however, there is quite a lot of research concerning payment schemes for participants in lab experiments. Many experimental economists most probably would insist that monetary risk is required in order to obtain valid conclusions about economic behavior. Payments based on the participants' performance are usually intended to provide incentives for rational – or at least well considered – decision making. On the other hand, there is evidence that monetary incentives do not necessarily increase performance [3]. All in all, we consider studying the impact of different incentive schemes on the prediction accuracy of markets an open and interesting question. We thus conducted a field experiment to analyze several monetary incentive schemes that could for instance be used in internal prediction markets for company-specific predictions.

3. EXPERIMENTAL SETUP

In this section we describe the setup of the field experiment we conducted during the FIFA World Cup 2006 in Germany. Firstly, we present the basic setup. Secondly, we elaborate on the three payment schemes we studied in our field experiment and explain why we chose these three incentive schemes. Thirdly, we discuss our expected results for this experiment.

3.1 Basic Setup

In our field experiment we were operating 20 prediction markets for the last 20 matches of the FIFA World Cup 2006. As assets we traded the possible outcomes of all the matches. There were three possible outcomes for every match – either team A won or team B won or there was a draw after the second half. We introduced the third asset ("draw") although there were no draws possible in the tournament. The reason was that we did not want to consider penalty shootouts because we considered their outcome more or less unpredictable. The asset corresponding to the events that actually occurred during the World Cup was valued at 100 currency unit after the match; the other two assets were worthless. Thus, traders could buy and sell so called unit portfolios comprising the three assets at 100 currency units at any time.

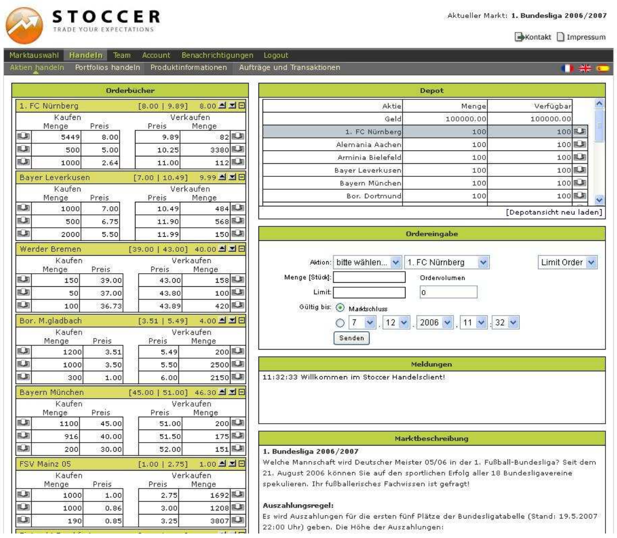

Figure 1. Web interface of the STOCCER trading platform

In total, 60 undergraduate students from the University of Karlsruhe, Germany, were taking part in our field experiment in June and July 2006. All the markets opened about two days before the corresponding match and closed at the end of the match. As a trading platform we used the system that is currently available at www.stoccer.com. A screenshot of the web interface is depicted in Figure 1. For more information on the system itself please refer to [6].

3.2 Incentive Schemes

We divided the 60 students randomly into three groups of 20 students each. At the end of the FIFA World Cup the users were paid according to their group's incentive scheme. We can thus study the impact of three different monetary incentives by comparing the prediction accuracy of the three groups of users, corresponding to three treatments with different incentive schemes. The subjects of the first group were paid a fixed amount of 50 Euro (from now on referred to as *FP*). In the second group, individuals were paid according to their ordinal rank (rank-order tournament, *RO*). The user ranked first was paid 500 Euro, the second 300 Euro and the third 200 Euro. All the other users in this group did not receive any payment at all. This also results in an average payment of 50 Euro per person. To subjects in the third group we promised what we called a performance-compatible payment, also with an average amount of 50 Euro (*DV*). Performance-compatible means that the payment linearly depended on the users' deposit value in the prediction market (deposit value divided by 10.000) and was therefore directly influenced by every transaction a user conducted.

We chose these three incentive schemes because we think they are somewhat related – although they are not the same – to incentives that we can nowadays typically observe in prediction markets, namely markets without any payment, real-money markets, and markets with rank-order tournaments. Comparing these three different monetary incentives is also of interest for operators of internal markets for company-specific predictions since companies are usually willing to reward their employees' effort. In this case, the question arises which monetary incentive scheme is the most suitable.

For every group we ran the 20 separate markets on 20 soccer matches that were described in Section 3.1. Since we did not want to pay students that were not trading at all we imposed a relatively small minimum trading volume per week on all of the users. Especially in case of the first group with the fixed payment we were worried that the students might otherwise consider not to trade at all.

3.3 Expected Results

Before conducting our field experiment we expected the third group with the performance-compatible payment to be the best and the first group with a fixed payment to be the worst in terms of prediction accuracy. In the following we explain the intuition behind these expectations.

For members of the first group, there exists no extrinsic motivation to reveal their expectations or to be among top performers of the group. In addition, there is no incentive for them to trade more than the minimum required trading volume per week. Members of the third group, on the other hand, receive a performance-compatible payment, meaning that every transaction directly influences their payment. Traders should consequently be motivated and try their best. Besides, traders don't want to loose money and will therefore consider very carefully what and how to trade. In short, traders with the incentive scheme *DV* have to "put their money where their mouth is" [4]. For the second group we expected a result somewhere in between the other two groups. On the one hand, traders have a strong incentive to be among top 3 traders of their group because they will not receive any payment otherwise. This should lead to a rather high trading activity. On the other hand, the rank-order tournament provides an incentive to take higher risk compared to traders e.g. in *DV*. Also, traders might start betting on unlikely events because they consider this the best or maybe even only way to outperform their competitors from the same group. For this reason, we expected that the inventive scheme *DV* would outperform *RO*.

4. RESULTS

In this section we will now discuss the – at first sight – probably somewhat surprising results from our field experiment. We will first compare the distribution of asset prices in the three treatments before discussing the impact of the three incentive schemes on the prediction accuracy.

4.1 Market Prices

In total, every group traded 60 assets in 20 different markets (three assets per market). In Figure 2 we can see how many assets were traded within a certain price range in each of the three treatments. The very first column for example means that 32% of the assets were traded at prices between 0 and 20 virtual currency units in the first treatment with a fixed payment.

When comparing the three treatments we can observe that a relatively high number of assets are traded at prices between 60 and 100 currency units in the second treatment. This is exactly what we expected because people are obviously willing to take the risk to buy assets even at rather high prices. Students in the third group with the performance-compatible payment, in contrast, do not trade any asset at a price between 80 and 100 currency units and almost not asset in the range from 60 to 80. Obviously, traders with *DV* are not willing to take the risk of buying assets at such high prices although there is no reason why their expectations should differ that much from the traders' expectations in the other two treatments.

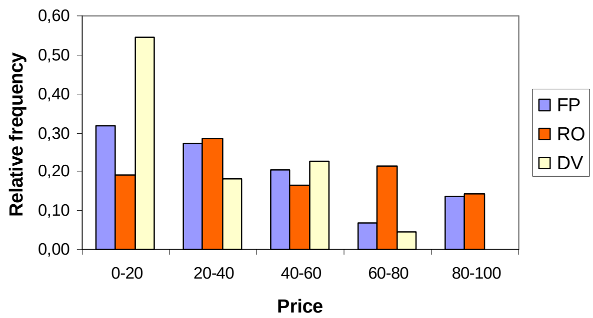

Figure 2. Distribution of asset prices in the three treatments

One again, "people are typically willing to pay less for almost anything if the money is real than if it is hypothetical" [9]. One explanation for this behavior of traders in the third treatment could be their risk aversion.

4.2 Prediction Accuracy

Overall, 35% of the assets with the highest share price out of the three assets per match actually corresponded to the observed outcome in case of the fixed payment and the average pre-game trading price of the asset corresponding to the outcome was 40.83 virtual currency units. In the rank-order tournament, the favorite outcome according to the asset prices actually occurred in 45% of the cases and the average pre-game trading price of the asset corresponding to the outcome was 51.65 currency units. Finally, in case of the performance-compatible payment, the favorite outcome according to the asset prices actually occurred in merely 20% of the cases and the average pre-game trading price of the asset corresponding to the outcome was 26.64 currency units. This means, when interpreting the asset prices as probabilities the third treatment predicted the outcome of a match worse than randomly drawing one of the three possible events. This was indeed rather surprising to us, especially since especially the rank-order tournament seems to work quite well.

However, in Section 4.1 we have already learned that asset prices seemed to be rather small in case of the performance-compatible payment. This can also be seen when calculating the sum of the three asset prices corresponding to the three possible outcomes of a match. These prices should sum up to about 100 virtual currency units since the probability that one of the three events occurs is 100%. In case of the performance-related incentive scheme the average price of such a so called portfolio is only 53.30 virtual currency units while it is indeed very close to 100 in the other two treatments.

To analyze the correlation between asset prices and outcome frequency in more detail, we sorted the data into buckets by assigning all of the assets to one of five price ranges according to their pre-game trading price. The size of the circles and triangles indicates how many assets prices fell into the price range. The larger the circle or triangle is, the more assets were assigned to this bucket. Figure 3 plots the relative frequency of outcome against the prices observed before the match started.

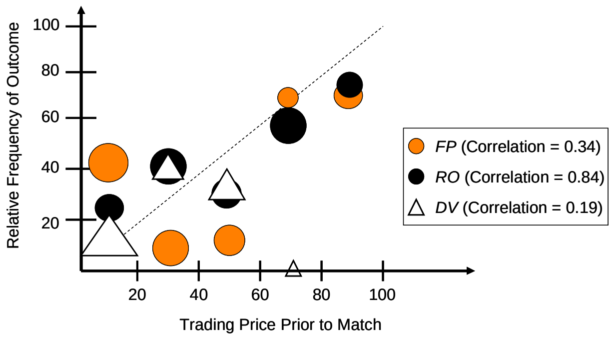

Figure 3. Market forecast probability and actual probability

For the rank-order tournament (black circles) the correlation coefficient is 0.84, while it is only 0.34 for the fixed payment and with 0.19 even worse for the performance-compatible incentive scheme. Thus, the prediction accuracy is – in contrast to our expected results – quite poor in the third treatment *DV*. Somewhat surprisingly, the rank-order tournament outperforms the other two incentive schemes.

As we have already mentioned earlier, on average the sum of the three asset prices corresponding to the three possible outcomes of a match was only 53.30 virtual currency units in case of *DV*. This might explain why the prediction accuracy is quite poor in case of this incentive scheme. To analyze this in more detail we divided all the asset prices by the average price of a portfolio and then once more plotted the relative frequency of outcome against the prices observed before the match started. Nevertheless, the rank-order tournament still performs much better than the performance-compatible incentive scheme.

4.3 Discussion of our Results

We can now only speculate about possible reasons for this result. Besides extrinsic motivation traders might also be intrinsically motivated. This could also help to explain why even the fixed payment scheme seems to work to some extent. However, we think that the risk aversion of the traders is most likely the main reason for our results. We conducted a lottery choice experiment as known from Holt and Laury [5] in order to measure the traders' degree of risk aversion before we started our field experiment. The choices involved large cash prizes that were paid to the participants. Nearly 75% of the subjects exhibit risk aversion.

In case of the fixed payment, traders can neither win nor loose money, so they just play for fun. Moreover, traders will take quite a lot of risk in the rank-order tournament because they have to be among the top performers within their group to receive the relatively high payment. Thus, the incentives over-ride risk aversion. Only in the third treatment, the performance-compatible incentive scheme, traders receive an endowment of 50 Euro and could potentially loose money with every transaction they make. As a result, buyers are obviously very careful and not willing to spend too much money on any asset. But why are sellers willing to give up assets at prices below their average worth? Well, users have to trade in order to reach the minimum transaction volume. Once sellers have started to partially sell their unit portfolios they are probably willing to sell at rather low prices to avoid the risk of holding shares of an event that does in the end not occur. Asset prices are thus much lower than in case of the other two incentive schemes. Maybe there would be almost no transactions if traders would not have to achieve the minimum transaction volume.

5. SUMMARY

In this paper, we have analyzed the impact of various incentive schemes on the accuracy of prediction markets. The results from our field experiment show that despite our first intuition performance-compatible payment schemes seem to perform worse than fixed payments and the rank-order tournament. Due to the risk aversion of traders, the competitive environment in case of the rank-order tournament seems to lead to the best results.

But what are the implications for designing future prediction markets? Well, out of the three incentive schemes we looked at one should choose the rank-order tournament when e.g. setting up an internal market for company-specific predictions where employees want to be paid for trading. We also argued in this paper that performance-compatible payment schemes are somewhat similar to real-money markets. But can we now draw the conclusion that play-money markets will outperform real-money markets although the latter raise numerous legal and technical difficulties? We would rather be careful when answering this question based on our results because the situation might be somewhat different in prediction markets that are open to the public. In this case, there is a self-selection of traders and we would thus expect many risk-seeking traders in such a real-money market. In such a situation a performance-compatible payment scheme might produce much better predictions than in our field experiment.

6. ACKNOWLEDGMENTS

This work is based on research funded by the German Federal Ministry for Education and Research under grant number 01HQ0522 and by the German Research Foundation (DFG) within the scope of the Graduate School Information Management and Market Engineering (IME). The authors are responsible for the content of this publication.

7. REFERENCES

[1] Fama, E.F. Efficient capital markets: A review of theory and empirical work. *Journal of Finance*, *25*. 383-417.

[2] Forsythe, R., Nelson, F., Neumann, G. and Wright, J. Anatomy of an Experimental Political Stock Market. *American Economic Review*, *82*. 1142-1161.

[3] Gneezy, U. and Rustichini, A. Pay Enough Or Don'T Pay At All. *The Quarterly Journal of Economics*, *115* (3). 791-810.

[4] Hanson, R. Decision Markets. *IEEE Intelligent Systems*, *14* (3). 16-19.

[5] Holt, C.A. and Laury, S.K. Risk Aversion and Incentive Effects. *American Economic Review*, *92* (5). 1644-1655.

[6] Luckner, S., Kratzer, F. and Weinhardt, C., STOCCER - A Forecasting Market for the FIFA World Cup 2006. in *4th Workshop on e-Business (WeB 2005)*, (Las Vegas, USA, 2005).

[7] Luckner, S., Weinhardt, C. and Studer, R. Predictive Power of Markets: A Comparison of Two Sports Forecasting Exchanges. in Dreier, T., Studer, R. and Weinhardt, C. eds. *Information Management and Market Engineering*, Universitätsverlag Karlsruhe, Karlsruhe, 2006.

[8] Manski, C.F. Interpreting the Predictions of Prediction Markets. *Economics Letters*, *91* (3). 425-429.

[9] Read, D. Monetary incentives, what are they good for? *Journal of Economic Methodology*, *12* (2). 265-276.

[10] Rosenbloom, E.S. and Notz, W.W. Statistical Tests of Real-Money versus Play-Money Prediction Markets. *Electronic Markets - The International Journal*, *16* (1).

[11] Servan-Schreiber, E., Wolfers, J., Pennock, D. and Galebach, B. Prediction Markets: Does Money Matter? *Electronic Markets - The International Journal*, *14* (13).

[12] Spann, M. and Skiera, B. Einsatzmöglichkeiten virtueller Börsen in der Marktforschung. *Zeitschrift für Betriebswirtschaft (ZfB)*, *74 (EH2)*. 25-48.

[13] Spann, M. and Skiera, B. Internet-Based Virtual Stock Markets for Business Forecasting. *Management Science*, *49*. 1310-1326.