Group Decis Negot
DOI 10.1007/s10726-017-9531-0

Manipulation in Conditional Decision Markets

Florian Teschner[1] · **David Rothschild**[2] · **Henner Gimpel**[3]

Abstract Conditional decision markets concurrently predict the future and decide on it. These markets price the impact of decisions, conditional on them being executed. After the markets close, a principal decides which decisions are executed based on the prices in the markets. As some decisions are not executed, the respective outcome cannot be observed, and the markets predicting the impact of non-executed decisions are void. This allows ex-post costless manipulation of such markets. We conduct two versions of an online experiment to explore scenarios in which a principal runs conditional decision markets to inform her choice among a set of a risky alternatives. We find that the level of manipulation depends on the simplicity of the market setting. When a trader is alone, has the power to move prices far enough, and the decision is deterministically tied to market prices or a very high correlation between prices and decision is implied, *only* then manipulation occurs. As soon as another trader is present to add risk to manipulation, manipulation is eliminated. Our results contrast theoretical work on conditional decision markets in two ways: First, our results suggest

Electronic supplementary material The online version of this article (doi:10.1007/s10726-017-9531-0) contains supplementary material, which is available to authorized users.

✉ Florian Teschner
florian.teschner@kit.edu

David Rothschild
davidmr@microsoft.com

Henner Gimpel
henner.gimpel@fim-rc.de

1 Karlsruhe Institute of Technology (KIT), Englerstr. 14, 76131 Karlsruhe, Germany

2 Microsoft Research, 1290 Avenue of the Americas, New York, NY, USA

3 University of Augsburg, Universitaetsstr. 12, 86135 Augsburg, Germany

Published online: 22 March 2017

that manipulation may not be as meaningful an issue. Second, probabilistic decision rules are used to add risk to manipulation; when manipulation is not a meaningful issue, deterministic decisions provide the better decision with less noise. To the best of our knowledge, this is the first experimental analysis isolating the effects of the conditional nature of decision markets.

Keywords Market design · Market manipulation · Conditional markets · Prediction markets · Decision markets

JEL Classification C9 · D8 · G1

1 Introduction

Modern information systems increasingly facilitate organizations to engage in participatory processes that harness collective intelligence by running prediction markets, for example. Prediction markets incentivize experts to reveal information and prove remarkably potent and robust as information aggregation mechanisms; frequently, they outperform other information aggregation mechanisms (Berg et al. 2008; Ledyard et al. 2009; Teschner et al. 2011; Bennouri et al. 2011). Canonically, markets have contracts that are worth \$1 if true and \$0 if false; with enough investors maximizing their return within the market, the market price is highly correlated with the probability of the outcome occurring (Wolfers and Zitzewitz 2006). Hence, many organizations endorse prediction markets—in a survey among 2609 entrepreneurs and employees, 9% reported to use prediction markets (Bughin et al. 2013). Corporate prediction markets are, for example, used to forecast product development performance, company or industry news, demand, and sales. While the crowd is used to gather and aggregate information and make predictions about the future; the final decision typically remains with a principal, e.g., the management board.

Prediction markets used as decision making tool are termed conditional decision markets or just decision markets (Hanson 1999; Berg and Rietz 2003). Such markets have a key difference to prediction markets: they both predict and decide the future (Chen et al. 2011). In decision markets, predictions need to be conditional on the decision being executed. Some decisions are not executed and, hence, the respective outcome cannot be observed. Markets for the non-executed decisions are void (Othman and Sandholm 2010; Chen et al. 2011). Behavior in markets that are void has no financial impact for the respective expert. Thus, decision markets allow for ex-post costless manipulation. Othman and Sandholm (2010) prove that experts can benefit from manipulating such conditional decision markets when the principal decides deterministically based on the markets' prediction.

We consider settings where the principal must make a choice over possible decisions to execute, where the possible decisions will impact the probability of achieving a desired outcome. The principal has access to experts who possess private information on the probability of the outcome conditional on the possible decisions. An illustrative example based on Chen et al. (2011); consider a project manager deciding whether to hire developer A or B for her team. The objective is to finish the project on time, which might be unequally likely for the two candidates. Assume the manager has access to knowledgeable experts, e.g., former colleagues of the two developers. To inform her decision, she invites the experts to participate in two decision markets: one predicting the likelihood to finish in time should developer A be hired, the other should B be hired. Assume developer B is hired. Then the manager can observe the development of the project and reward experts based on their accuracy in predicting the likelihood given B. She will never know what would have happened with developer A. Hence, market A is void. Now assume, sometime before the markets are closed, the markets' prices are \$0.60 for finishing in time given the recruitment of A and \$0.80 given B (correlating to 60% for A and 80% for B) and an expert has the belief that the efficient prices would be \$0.70 and \$0.80 (correlating to 70% for A and 80% for B). In either case—following the current market prediction or the single expert—B should be hired. This decision does, however, not yield a financial benefit for the expert in question, as he cannot improve the prediction given the recruitment of B. On the other hand, he can correct market A to \$0.70 and manipulate market B to below \$0.70. With this, the manager will hire A, the expert expects a positive profit from trading in A, and manipulation in market B is costless, as it becomes void. Such behavior is termed "manipulation", because it leads to a decision that does not confirm the expert's beliefs.

Two features are required for this manipulation incentive to occur with decision-agnostic experts. First, markets need to be conditional (i.e., not all markets are executed) so that there is the possibility that manipulation can be costless. Second, the price of the markets needs to have an effect on which market is executed and which is void. Othman and Sandholm (2010) and Chen et al. (2011), among others, study such conditional decision market settings theoretically for fully rational, risk-neutral, self-interested, expected-utility-maximizing experts.

To complement this theoretical perspective, we study conditional decision markets empirically in two online experiments. We create a decision task, recruit and instruct participants, train them on our market, and weed out potential participants who do not understand the market. We then turn participants who proved having understood the experimental setup into experts by providing them information and allowing them to trade in decision markets for eight rounds. In Experiment 1, participants solely interact with a market maker and in Experiment 2 they interact with a market maker and one fellow trader. Each participant is randomly assigned to one of four treatments where treatments differ by the decision rule applied. With this, we study how and when participants manipulate the markets depending on to experimentally controlled factors, the changing quantity of traders and the decision rules. Our experiment sheds light on the extent and influence of manipulation on information aggregation and decision quality. The contribution of our work is that we study informed manipulators in very stylized conditional decision market settings and provide the first empirical evidence that the issue is not substantial.

Our experimental results show that manipulation impairs information aggregation, however, only to a very limited extent. The extent of manipulation depends on the decision rule—manipulation disappears as the correlation between the market prices and the decision disappears. While manipulation in conditional decision markets occurs in highly stylized settings, our data suggest that it does not in the commonly used practice of running decision markets. As soon as another trader is present to eliminate the guarantee of last move, manipulation is eliminated. While theoretically a probabilistic decision rule should preclude manipulation, we find no evidence for this in our data. Rather on the contrary, a probabilistic decision rule leads to more incorrect decisions than a deterministic one. We note the difficulties in generalizing experimental work to the real world, but understanding those considerations these results add to the knowledge base of market design research; designers can draw on these insights and apply them to future corporate decision markets.

2 Related Work

Scoring rules are well known as incentive compatible mechanism to elicit a single expert's belief in the probability of an outcome (McCarthy 1956; Savage 1971; Gneiting and Raftery 2007). In recent years the application of this traditional mechanism has been extended by adding two levels of strategic complexity.

1. *Multiple experts* Hanson (2003, 2007) and Pennock and Sami (2007) *extended the notion of single-expert scoring rules to scoring rules sequentially shared among multiple experts.*
2. *Decisions* Othman and Sandholm (2010) *formalized the strategic interaction of experts with the principal when the event upon which predictions are conditioned is not random but its realization is decided by the principal.*

These two levels of strategic complexity lead to overall four settings—each of which has been studied theoretically.

Single expert, random event (Scoring rule) Facing a proper scoring rule, a single fully rational self-interested risk neutral expert maximizes expected utility by truthfully revealing his belief (Chen et al. 2011; see Hanson 2003, p. 109 for a brief review on scoring rules). The same holds true for a set of parallel, unconditional elicitations. In a more complex setting, the principal might be interested in eliciting the probability of an outcome conditional on an event. With a random event (not to be influenced by either the principal or the expert), this can be reduced to eliciting the beliefs regarding the joint probability distribution of events and outcomes—again, scoring rules are incentive compatible.

Multiple experts, random event (Market scoring rule) A scoring rule sequentially shared by multiple experts is myopic incentive compatible (Hanson 2003). An expert interacting with the scoring rule only once or for the last time maximizes expected utility by truthfully revealing his belief. An expert with foresight and multiple interactions with the sequentially shared scoring rule can, however, benefit from delaying his revelation or bluffing his fellow experts (Dimitrov and Sami 2008; Chen et al. 2010). An alternative interpretation of a sequentially shared scoring rule is seeing it as prediction market with an automated market maker (Hanson 2003). With this, market scoring rules tie into the large body of literature on prediction markets, e.g. based on continuous double auctions or pari-mutuel betting.

Single expert, decision (Conditional scoring rules) Othman and Sandholm (2010) were the first to formalize the incentive problem when a single expert knows that the principal will decide based on the expert's prediction which in turn affects which of the scoring rules is payoff relevant for the expert. The principal's most natural decision rule is deciding in favor of the alternative where the expert predicts the highest likelihood of achieving the outcome desired by the principal. Othman and Sandholm (2010) call this decision rule the 'max decision rule' and prove that no symmetric scoring rule nor any asymmetric scoring rule from literature is incentive compatible.

Randomness can result in incentive-compatibility, as e.g. seen in voting mechanisms (Gibbard 1977; Wagman and Conitzer 2008). Chen and Kash (2011) apply this idea to single expert with conditional scoring rules and prove that with a probabilistic decision rule incentive compatibility is restored. This comes, however, at the cost of the principal knowingly taking a sub-optimal decision with positive probability.

Multiple experts, decision (Conditional decision markets) Incentive compatibility of single expert, random event scoring rules is challenged by both the existence of multiple experts and the payoff being conditional on the principal's decision. Not surprisingly, with a deterministic decision rule, no (known) market scoring rule is incentive compatible, not even myopic incentive compatible; an expert rather benefits from exaggerating the success probability of a suboptimal decision (Othman and Sandholm 2010; Chen and Kash 2011). The intuition: as markets predicting based on counterfactuals are void, experts deterministically know that manipulation in markets which are void is costless. They use this costless manipulation to steer the decision in a direction that increases their expected gains from trade. Myopic incentive-compatibility can, however, theoretically be achieved with a probabilistic decision rule that makes experts indifferent between their gains from manipulation and the expected costs of manipulation (Chen et al. 2011).

In summary theory suggests that conditional decision markets are strategically complex. First, experts can benefit within a single market by delaying trade or bluffing fellow traders. There is no full characterization of equilibrium behavior by fully rational risk neutral decision-agnostic experts (not to mention a relaxation of these assumptions). Second, with a deterministic decision rule they can benefit from strategically manipulating the principal's decision. A probabilistic decision rule annuls this incentive and should, consequently, result in the same accuracy of prediction as market scoring rules either predicting conditional on random events or unconditional market scoring rules. Thus, we study four settings that cover all of these key alternative scenarios, we denoted them as: DETERMINISTIC, PROBABILISTIC, RANDOM, and UNCONDITIONAL.

All of the above holds for decision-agnostic experts. Recently, several papers analyzed experts having a vested outside interest in the principal's decision (Dimitrov and Sami 2010; Gimpel and Teschner 2013a, b). Empirical evidence on the issue of users maximizing their outside vested interest relative to the market's price is mixed. Hanson et al. (2006) find that manipulators are unable to distort price accuracy. Rhode and Strumpf (2006) show that markets with many users confined to limited amounts of wealth are very hard to manipulate. In addition, Rothschild and Sethi (2013), argue that principals do not rely just on the final price, but the trading as well. Hence, it is relatively easy to identify manipulation strategies. On the contrary, Deck et al. (2013) present experimental data on manipulation being successful and market observers being tricked by manipulation. Similarly, Gimpel and Teschner (2013b) present data on manipulation in experimental conditional decision markets where experts have an outside incentive to manipulate the decision. The present paper is, however, restricted to decision-agnostic experts to isolate the effect of conditional markets.

More generally, on the empirical side, research on prediction markets suggests that prediction markets generally aggregate dispersed information well; typically better than other information aggregation and prediction mechanisms (Berg et al. 2008; Ledyard et al. 2009; Teschner et al. 2011; Bennouri et al. 2011). Market scoring rules tend to be easy to use and produce reliable predictions (Healy et al. 2010; Jian and Sami 2012). Overall, empirical evidence suggests that prediction markets are by and large robust to manipulation (Rhode and Strumpf 2006; Deck et al. 2013). To the best of our knowledge, however, behavior in conditional decision markets has not yet been studied empirically.

3 General Design and Procedures

The experiments compare information aggregation and market outcomes in stylized conditional decision market settings. There are two binary lotteries represented by two urns, A and B, holding 100 balls each. Per urn, 32 or 68 of these are black, the others white. A principal—who will draw one ball from one of the urns—is interested in drawing a black ball. She decides which urn to draw from, the draw itself is random. The principal does not know the number of black balls in either of the urns. To gain information prior to deciding, she runs two parallel prediction markets for experts to share their private information: One market for urn A and one for B. The market price is assumed to reflect the aggregate prediction of the probability to draw a black ball from the respective urn. Experts are financially compensated based on their trading performance. Recall the hiring decision from Sect. 1: the two urns represent the two potential employees. Drawing a black ball stands for the desired outcome to get the project completed on time. The executive is the principal and employees participating in the market are the “experts”.

Each experiment is a series of 8 payoff-relevant periods. Each period follows 3 phases.

1. *Private information:* The number of black balls per urn is determined and participants receive fully revealing private information. Thus, participants become experts via private information (like in the experiments by Oprea et al. (2007), Healy et al. (2010), Jian and Sami (2012), Deck et al. (2013) and many others before that). Further, the true state of each urn remains visible to the participants during the trading period.
2. *Prediction market:* Participants can buy and sell virtual stocks in two parallel markets. The market uses a logarithmic market scoring rule (Hanson 2003). The value of the stocks is linked to the color of the ball that will be drawn. The final market price is used by the decision maker as the best predictor for the number

of balls in the respective urn. Bennouri et al. (2011), Jian and Sami (2012), and others use similar approaches.

3. *Decision:* After the markets close the decision maker picks an urn and a ball is drawn (in treatment UNCONDITIONAL, two balls are drawn, one from either urn). Across the two experiments, we consider four different treatments that correspond to decision strategies. In UNCONDITIONAL, there is no decision but the principal draws from both urns. In the other three treatments, the principal decides which urn to draw from. In RANDOM, she relies on hidden or unobserved information; thus, we set this a-priori probability to 50% for either urn, because that is what it is to the experts, who are not influencing the decision. In DETERMINISTIC, she chooses the urn where the markets predict the higher likelihood of drawing a black ball, (i.e. the urn with the higher final market price). In the case of tied prices, the choice is random. In PROBABILISTIC, the principal uses a logit decision function: each choice has positive probability but she is more likely to choose the urn with the higher market price; the likelihood increases with the price difference. The exact form is defined in the Appendix in Electronic supplementary material. Markets are conditional on the principal's choice (Berg and Rietz 2003); for the urn chosen by the principal, a ball is drawn and its color determines the experts' compensation for trading. The other market is voided. Experiment 1 includes all four treatments, while Experiment 2 uses just RANDOM and DETERMINISTIC; we explain this difference in more detail below.

The main substantial difference between the two experiments, noted in more detail below, is that in Experiment 1 we do not note the existence or non-existence of another trader (there is none), but in Experiment 2 we explicitly note the existence of one other trader. The default state among Mechanical Turk users is that they perform their work alone, without interaction with other Mechanical Turk workers. In Experiment 1, we do not address the non-existence of another trader as we do not want the participants to overthink the implications of something that is default assumption. In Experiment 2 we explicitly inform participants on the deviation from default.

When participants arrive at the market there is an initial price the market maker is willing to trade. The participants are never told if the starting position was determined through earlier trading or the market maker has determined it or if it is random. The initial pricing was randomized over a pre-determined set of prices to ensure that the participants encountered a meaningful selection of initial scenarios which would induce varying theoretically maximizing trading strategies. The values were randomized between rounds and right and left in the trading platform. The eight starting price scenarios are noted in the Appendix in Electronic supplementary material.

We recruited participants via the online labor market Mechanical Turk. Recruitment was restricted to workers registered as residing in the US. Each participant took part once in exactly one treatment in one experiment (between-subject design). As participants dropped out or failed our pre-experiment survey the numbers are not fully balanced over the treatments. The experiment was run with a custom-made web application on Mechanical Turk. From an organizational and technical perspective we followed the guideline of Mason and Suri 2012. We add more detail on Mechanical Turk users in the Appendix in Electronic supplementary material. Also, we prop-

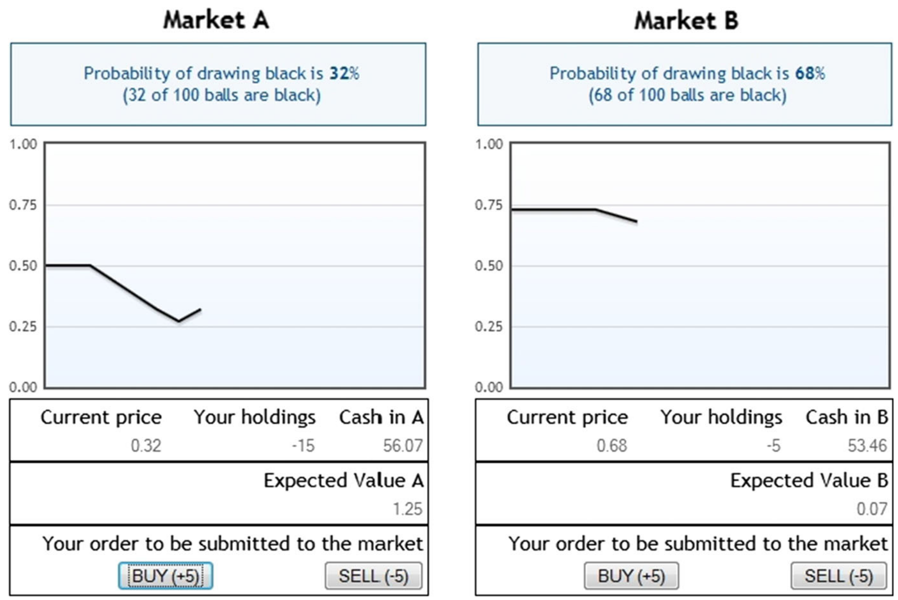

Fig. 1 Screenshot from the trading interface in Experiments 1 and 2

erly incentivized participation and performance. Each participant could take part at their own pace without a time limit. Sessions lasted around 15 minutes. Payments were linked to individual performance in the experiment; the average payment was \$2.49, with a variance of \$0.26 in Experiment 1 and \$2.42, with a variance \$0.28, in Experiment 2. (Instructions for participants are attached in the Appendix in Electronic supplementary material). See Fig. 1 for a screenshot of the trading interface for phase 2, the prediction market.

Participants have full knowledge of the rules of the game including the decision rules. We wrote the instructions to ensure that we had all of the correct technical terminology, but also included more vernacular wording for the participants to understand. Prior to trading, we test all potential participants with key questions about the markets and rules, and drop anyone who cannot get perfect scores; we drop nearly half of all potential participants, consistent with both experiments, through this test.

Contrary to the forecasters in the experiments by Oprea et al. (2007) and Deck et al. (2013), in our experiments the principal is automated and her decision rule depends on the experimental treatment. In Experiment 1 we compare an UNCONDITIONAL situation with three treatments—termed RANDOM, PROBABILISTIC, and DETERMINISTIC—in a between subject design. RANDOM corresponds to the single expert, random event (scoring rule) scenario (cf. Sect. 2), PROBABILISTIC and DETERMINISTIC are variants of the single expert, decision (conditional scoring rules) scenario. In Experiment 2 we pare this down to just RANDOM and DETERMINISTIC.

The RANDOM and UNCONDITIONAL determination serve as benchmarks. UNCONDITIONAL does not provide incentives for manipulation, as there is no decision to be manipulated. The same holds for RANDOM, which has a decision but the decision is not to be manipulated by strategic trading.

The PROBABILISTIC determination mimics the current common use of conditional decision markets in practice. The markets' information is taken into account but the principal has discretion to decide otherwise. DETERMINISTIC is an alternative where the principal pre-commits to following the markets' suggestion. In DETERMINISTIC, experts can sometimes gain more by trading in the market with lower likelihood of approval. Thus, depending on an expert's risk aversion and the potential gain in each market, she will prefer either a choice of urn A or B. The same hold for PROBABILISITC but, following theory, to a lesser extend as expected costs of manipulating become strictly positive.

In the PROBABILISTIC and DETERMINISTIC treatments, experts have an incentive to manipulate the market price to willingly mislead the principal to increase gains from trade. In the DETERMINISTIC treatment, there is no risk of the inefficient market being chosen, if the final price is less than the other market, but in the PROBABILISTIC treatment, this potentially costly risk is unknown to the participants.

Participants' behaviors in the UNCONDITIONAL and RANDOM treatments should be about equal and serve as empirical benchmark for manipulation-free behavior. In DETERMINISTIC one should observe manipulation and, thus, some level of incorrect decisions. PROBABILISTIC should theoretically feature less manipulation—whether this holds true and whether any manipulation remains are questions to be explored empirically. These decision rules are the same irrespective of whether the principal is assumed to be risk neutral or risk averse.

There are several factors that should influence the amount of manipulation. First, we weed out nearly half of the potential participants, with our pre-experiment survey, to ensure that the participants really understood the decision rule, the payout structure, and the potential to benefit from manipulation in treatments PROBABLISITC and DETERMINISTIC. Second, we are very explicit in the introduction what the expected payout is by market and how a participant maximizes that in any given market. Third, to ensure the participants can see which market would payout more, we do the math for them and show the expected payout in each market explicitly on the trading wizard. Fourth, we provide an exact underlying probability (i.e., composition of each urn), where any uncertainty of the probability would decrease the incentive for manipulation, because the participant could not be sure they getting a higher return for their risk. Fifth, in the results below we drop any participants who did not trade, as a minimum test to ensure the results are not overwhelmed by noise from lazy participants. Sixth, the return from manipulation exceeded the return from honest revelation on average by a factor of 5. As we paid relative to performance per treatment there was a significant incentive to manipulate.

4 Experiment 1: One Trader Results

As noted above, there are two distinctions between this experiment and Experiment 2: traders and treatments. First, to clearly control for the incentives and beliefs of a par-

Table 1 Participant characteristics for Experiment 1

	UNCONDITIONAL	RANDOM	PROBABILISTIC	DETERMINISTIC
n	67	85	134	112
Age (years)	34.9	33.1	34.2	31.8
Share female	43%	40%	34%	42%
Duration (min)	15.0	13.3	15.8	14.5
# Trades (per round, mean/median)	11.5/9	9.3/6	10.3/7	10.5/8
# Trades (per round: min/max/var)	1/118/162.8	1/182/158.2	1/199/154.5	1/98/100.6

ticipant, there is no interaction with other traders (unlike, e.g., Healy et al. 2010 or Jian and Sami 2012). Second, we use all four decision rules as treatments in Experiment 1.

We start by describing the basic statistics on participants and then turn to analyzing manipulation and its effects. Manipulation is assessed via absolute differences from market prices to the true underlying state of nature. Recall that participants know the true underlying value, so are intentionally providing false information to the principal. Effects of manipulation are assessed by analyzing how a principle would decide based on market information and in how far this decision coincides with her preferred decision she would make would she have perfect information on the state of nature. We call deviations from the preferred decision given private information, to the decision given market information, as “binary incorrectness”. In this, we disentangle binary incorrectness resulting from misleading market prices on the one hand and from probabilistic elements in the decision rule on the other hand.

Descriptive: Due to random assignment, participant characteristics are balanced across treatments. Age and gender are very similar. There is no significant difference between the treatments in duration (which was not limited for the participant) and number of trades per round. Table 1 has the details. Each participant participated in eight sets of two markets, all with the same treatment. Again, we have dropped participants who did not trade. While introduction into the treatments is random, we did oversample PROBABILISTIC and DETERMINISTIC.

We analyze the information aggregation in our stylized market setting by examining the pricing error per treatment. As we know the underlying probability of the outcomes, we can ex-post calculate the absolute error, (i.e., the absolute difference between the last market price and correct price given the true probability). We break up the results into two different groups, first situations where it would be optimal for the participant to manipulate if there was a deterministic decision rule (i.e., there is a higher expected value from the market with the lower underlying probability) and second situations where it would not be optimal for the participant to manipulate. This breakdown assumes purely financially motivated experts without innate preference for truth telling or manipulation. By design, the breakdown between groups is almost perfect with 1,548 market pairs in situations where manipulation is efficient and 1,544 market pairs where it is not. In a pair of markets, manipulation might affect either of the markets or both markets. In order to detail the treatment differences we use linear

Table 2 OLS regression on absolute errors by treatment for Experiment 1

	Manipulative situations		Non-manipulative situations	
	Estimate	Std. Error	Estimate	Std. Error
(Intercept)	33.38***	2.29	29.72***	2.18
RANDOM	−2.02	2.98	−1.61	2.66
DETERMINISTIC	**5.74****	2.61	−1.25	2.50
PROBABILISTIC	**5.83****	2.58	0.63	2.33
Round	0.26	0.25	−0.04	0.22

** Significance at 1%, and *** significance at 0.1%. The manipulative situation has 1548 observations and a R^2 of 0.021. The non-manipulative situation has 1544 observations and a R^2 of 0.002. Standard errors are clustered on the respondent

regressions on the sum of absolute pricing errors in both markets in a pair. As baseline we use the UNCONDITIONAL treatment which has no decision rule. Table 2 gives the regression results.

As expected we find no significant difference between our two benchmark treatments RANDOM and UNCONDITIONAL for either manipulative or non-manipulative situations. Similarly, we find no significant difference between DETERMINISTIC and PROBABILISTIC for either manipulative or non-manipulative situations (the p-value is 0.95). Errors in both DETERMINISTIC and PROBABILISTIC are significantly higher than in UNCONDITIONAL for manipulative situations whereas no such difference exists for non-manipulative situations. This clearly indicates that participants in DETERMINISTIC and PROBABILISTIC acted differently and provide the principal prices that are not as close to the true probability when they have an incentive for manipulation. The effect size is, however, only small to small/medium—Cohen's d ranges from 0.24 to 0.33 for different pairs of UNDCONDITIONAL/RANDOM and DETERMINISTIC/PROBABILISTIC treatments. 0.2 is commonly considered a small effect and 0.5 a medium effect.

As a robustness check the magnitude and significance are very similar as regressions with interactions of the manipulation setting, if the errors are squared, and various attempts at controlling for individual-level errors. As a second robustness check we run an anova. We see that there is a significant effect of the treatment and on the trading error $[F(3, 1544) = 10.94, p < 0.1\%]$ for the states with manipulation incentives (The post-hoc results are displayed in the Appendix in Electronic supplementary material). Using the same analysis for the non-manipulative states we find no differences $[F(3, 1540) = 1.02, p = 0.38]$.

Market prices are the outcome of trading behavior. As an additional metric for manipulation, we consider trading behavior. Specifically, for each individual trade we classify whether it drives the price towards the correct value for that market or not. We then calculate the ratio of trades that participants do to drive the price towards the correct value and compare it across treatments. The lower the ratio, the more manipulation. Table A2 in the Appendix in Electronic supplementary material has the results. In the DETERMINISTIC/PROBABILISTIC treatments traders are more likely to go in the opposite direction of the true underlying value than the baseline. And,

Table 3 Logit regression on binary incorrectness by treatment and situation for Experiment 1

	Manipulative situations		Non-manipulative situations	
	Estimate	Std. Error	Estimate	Std. Error
(Intercept)	−1.75***	0.19	−1.67***	0.20
RANDOM	0.11	0.22	0.07	0.32
DETERMINISTIC	**0.43***	0.21	0.09	0.22
PROBABILISTIC	**0. 39***	0.21	0.10	0.21
Round	0.07**	0.03	0.01	0.03

* Significance at 5%, ** significance at 1%, and *** significance at 0.1%. The manipulative situation has 1548 observations and a Pseudo-R^2 of 0.010. The non-manipulative situation has 1544 observations and a Pseudo-R^2 of 0.000. Standard errors are clustered on the respondent

this manipulation pays off; participants in the DETERMINISTIC/PROBABILISTIC treatments make more money, on average, than the participants in UNDCONDITIONAL/RANDOM (see the Appendix in Electronic supplementary material for details).

There are treatment differences in the markets' information revelation. The next question is, whether these significant yet small/medium differences have a meaningful and significant impact on the principal's decision. Each time the principal faces a decision, she has a preferred decision (i.e., drawing from the urn with more black balls). Thus, we can measure the rate of incorrect decisions by treatment and situation. To analyze this, we first apply a deterministic decision rule to any pair of markets, irrespective of the treatment. Subsequently, we analyze the treatment decision rule itself applied to market prices. This two-step process disentangles two sources of incorrect decision: imprecise information and randomness induced by the decision rule itself.

First, we apply a deterministic decision rule to market prices from each treatment: Given the differences in information revelation by treatment and situation (cf. Table 2), one should expect differences in incorrect decisions. Logit regressions show that indeed statistically significant differences in binary incorrectness exists. The results in Table 3 show a statistically significant increase in the likelihood of incorrectness for the DETERMINISTIC and PROBABILISTIC treatments only. In the non-manipulative setting this increase does not exist. As for the absolute errors, again, there is not a statistically significant difference between the DETERMINISTIC and PROBABILISTIC treatments for either manipulative or non-manipulative situations (the p-value is 0.77).

In the manipulative situations, both the error and the binary outcome are meaningfully positively correlated with the round within the experiment. It is not significant in the error outcome, but it is in the binary outcome. Also, it is important to note here that we have eight rounds, so the effect has a non-negligible impact. Yet, we do not find a significant positive interaction with DETERMINISTIC and/or PROBABILISTIC treatments (data not shown). Thus, we cannot rule out that this is tied to some underlying round effect that is exogenous to learning about manipulation.

Table 4 demonstrates the percent of the time that the lower probability market has the higher final price per treatment (i.e., how often would the principal pick the

Table 4 Binary incorrectness by treatment for Experiment 1 with different decision rules

	UNCONDITIONAL	RANDOM (%)	PROBABILISTIC (%)	DETERMINISTIC (%)
Deterministic decision rule	17.4%	18.9	21.5	21.8
Treatment's decision rule	–	50.0	40.2	21.8

wrong market if she applied a deterministic decision rule). Unsurprisingly, this happens more often in PROBABILISTIC and DETERMINISITIC (Table 4, first row). For the comparison of UNDCONDITIONAL/RANDOM on the one side and DETERMINISTIC/PROBABILISTIC on the other side, we estimate Cohen's d based on odds ratios (Chinn 2000): it ranges from 0.09 to 0.15. By convention, this is a small effect. For manipulative settings only, the effect size ranges from 0.15 to 0.24, i.e. small. We thus conclude that in manipulative settings in the DETERMINISTIC and PROBABILISTIC treatments our experiment reliably creates and measures manipulation which does, however, only marginally worsen the information provided to the principal.

However, a decision maker running prediction markets is interested in the choices she has to make based on the design of the markets. Hence she is interested in the correctness after applying the ex-ante communicated decision rule. We applied the treatment's decision rule on our data we get 50% incorrect by design for RANDOM and 40.2% for PROBABILISTIC. It is easy to see that in this case, if she cannot run UNCONDITIONAL her best choice is a DETERMINISTIC rule. The reduction of manipulation suggested by theory under the PROBABILISTIC rule is—if it exists at all—out weighted by the incorrectness of applying the decision rule.

Of course, the PROBABILISTIC decision rule can vary from close to the RANDOM to close the DETERMINISTIC and one could argue whether our experiment featured the optimal probabilistic decision rule for the principal. Let us assume that the rule was closer to random than optimal; in that case, an excessively random decision rule did not substantially lower manipulation, as binary incorrectness is virtually indistinguishable between PROBABILISTIC and DETERMINISTIC. We would conclude that probabilistic elements do not solve the issue of manipulation. On the contrary, assume the rule was closer to deterministic than optimal; following theory, more randomness should result in less manipulation, i.e., moving the binary incorrectness with a deterministic rule from the observed 21.5% closer to 18.9%. Given that the probabilistic decision rule used in the experiment already distorts 21.5–40.2% incorrectness, it appears unlikely that more truthful behavior paired with more random decisions could bring incorrectness down to the 21.8% for DETERMINISTIC. In summary, our data suggest that a probabilistic decision rule does not outperform a deterministic rule.

There are two main findings of the first experiment. First, respondents do manipulate the prices in the DETERMINISTIC and PROBABILISTIC treatments. A select group of participants move the price of the higher probability outcome downward in order to ensure or raise the probability of the lower probability outcome market being chosen by the principal with the expectation of a higher return. Second, any benefits of having a probabilistic scoring rule, which could deter manipulation, are dominated by the cost of not choosing the outcome picked by the market, which is generally accurate.

5 Experiment 2: Two Trader Results

We play the same setting but now match people into pairs. So, rather than being the last play in the markets, the participants are now matched against a live partner. It is still possible to guarantee a higher expected profit with manipulation, but it is much harder for participants to both see and realize that strategy. Theoretically, we move from single expert scenarios tested in Experiment 1 to multiple expert scenarios in Experiment 2. Again, we compare both the random event and decision variants to test the different theoretical predictions outlined in Sect. 2.

The specific steps of the game are as identical as possible to Experiment 1 while accommodating the change to a two player setting. First, the instructions are just slightly updated to state that there is another player and a set amount of time (3 min for the first round to get used to the experiment and the interface, 2 min per round thereafter). Second, knowledge of this is inserted into the pre-test. Third, potential participants are placed in a waiting room until a pair arrives. Most players waited just a few seconds, as the game filled up quickly (half of matched pairs waited no time, as their match was in the waiting room). A few players towards the end waited the full possible time of 5 min without a match and were just paid for their time.

With the complexity and loss of power (we would have needed twice as many participants for the same number of completed markets) we decided to limit the experiment to two treatments. Since we saw no significant differences between the DETERMINISTIC and PROBABILISTIC treatments, we only keep the DETERMINISTIC. The DETERMINISTIC has a higher likelihood of manipulation, in theory, so again we want to allow for that. Similarly, we saw no significant differences between the UNCONDITIONAL and the RANDOM treatment, so we use the RANDOM treatment only.

The ability to manipulate is slightly different in this setting. While Experiment 1 had no clock, we provide a clock in this version to allow participants to manipulate right at the end with diminished concern of the other player countering their move. But, in some scenarios the other player could sabotage a manipulation strategy for

Table 5 Participant characteristics for Experiment 2

	RANDOM	DETERMINISTIC
n	92	94
Age (years)	30.7	30.8
Share female	43.1%	48.9%
# Trades (per round, mean/median)	14.2/10	14.8/11
# Trades (per round: min/max/var)	1/71/158.5	1/154/175.1

Duration is pre-set at 17 min for all participants in Experiment 2, as the experiment was timed, such that participants stayed synchronized throughout the session

Table 6 OLS Regression on absolute error for Experiment 2

	Manipulative situations		Non-manipulative situations	
	Estimate	Std. Error	Estimate	Std. Error
(Intercept)	39.84**	4.71	29.60**	3.68
DETERMINISTIC	−0.60	5.01	−2.07	4.43
Round	−0.05	0.67	0.80	0.74

** Significance at 1%. The manipulative situation has 283 observations and a R^2 of 0.000. The non-manipulative situation has 274 observations and a R^2 of 0.007. Standard errors are clustered on the trader pair (cohort)

their advantage. Imagine a situation where player 1 has driven the price of the lower probability market to its true value and driven the price of the higher probability market downward to ensure that the lower probability market would be chosen. This is risky, as the second player could drive the price in the higher probability market back up, reaping a high expected return and leaving player 1 with a lower expected return in the market that is ultimately chosen. Of course, player 1 can make the move on the higher probability market right at the end as the clock closes out trading, locking in the expected profit.

We use the same set of characteristics as in Table 1 in Table 5 to show that the users are similarly representative. Each participant was randomly paired in a cohort with one other participant and then participated in eight sets of two markets, all with the same treatment.

For the following analysis we selected only rounds in which both participants were active; results are robust when using a less strict selection criterion. We ran the regression with the RANDOM treatment as the baseline and the dummy for DETERMINISTIC was slightly negative and insignificant, for both the manipulative and non-manipulative situations (Table 6). Since there is no benefit for manipulation with a random decision method, we believe this is a strong demonstration of no manipulation in the DETERMINISTIC treatment.[1]

With no meaningful difference in the error it is not surprising we find no significant difference in the binary answer. As robustness check, we again run an anova finding no significant treatment effects in both settings. Table 7 shows the results for the logit regression on whether the decision would be correct if it utilized a deterministic rule. There is no meaningful or significant coefficient in either the manipulative or non-manipulative situations on the DETERMINISTIC versus the RANDOM treatment.

There are two main findings of the second experiment. First, respondents no longer manipulate when you add a second player. Second, while Experiment 1 has some indication of the possibility of learning, in the more realistic Experiment 2 there is no learning effect in the manipulative situation that would indicate participants learning

[1] Comparing Tables 2 and 6, the absolute pricing error in the RANDOM treatment appears to be higher in Experiment 2 as compared to Experiment 1. We suspect that this might reflect the higher complexity and anxiety of having an opponent and a clock in Experiment 2.

Table 7 Logit regression on binary incorrectness by treatment for Experiment 2

	Manipulative situations		Non-manipulative situations	
	Estimate	Std. Error	Estimate	Std. Error
(Intercept)	−1.16**	0.30	−2.11**	0.39
DETERMINISTIC	0.18	0.29	−0.48	0.41
Round	0.00	0.06	0.17+	0.09

+ Significance at 10% and ** significance at 1%. The manipulative situation has 283 observations and a Pseudo-R^2 of 0.002. The non-manipulative situation has 274 observations and a Pseudo-R^2 of 0.040. Standard errors are clustered on the trader pair (cohort)

how to manipulate. This finding is robust to testing a range of further interactions not shown in the tables.

6 Conclusion

In this paper, we review the current literature on conditional decision markets and their antecedents. To understand the extent and influence of manipulation on information aggregation and decision quality, we study such markets experimentally. The contribution of our work is that we study informed manipulators in very stylized conditional decision market settings. Decision markets have been extensively analyzed in theory; we provide the first empirical evidence that the issue is not substantial.

In line with select previous work (e.g., Deck et al. 2013), our experimental results show that manipulation impairs information aggregation. The extent of manipulation depends on the decision rule—manipulation disappears as the correlation between the market prices and the decision disappears. While manipulation in decision markets occurs in highly stylized settings, it vanishes once the setting becomes more natural.

Specifically, we experimentally test the theoretical prediction that a probabilistic decision rule reduces manipulation and hence improves the decision maker's decision quality. We find the contrary, manipulation is not reduced and the probabilistic decision rule adds randomness to the decision taken. As for any experimental work, one has to be careful in generalizing the results as the details of the environment matter. The reward mechanism, different incentives in the field, the time for the experiments, the instructions, risk attitudes, the type of decision to be made—all of these play a role. Nevertheless, theoretical predictions on decision markets should apply in our experimental settings but our data only partially supports these theoretical predictions. Thus, despite the general limits to external validity, we believe that our experimental results shed some light on which theoretical predictions to follow when implementing decision markets.

We used different metrics on participants' behavior and market outcomes to test for manipulation. Strictly speaking, the metrics show that depending on the treatment and situation, participants deviate from truth-telling. Besides intentional manipulation, there might be other reasons for these deviations. Nevertheless, we refer to such deviations from truth-telling as "manipulation" throughout the paper as they are in line with theoretical predictions on where to expect manipulation and in which direction (buy/sell, price up/down) participants might want to manipulate.

In both experiments we design the platform to encourage manipulation; in the real world it will be more difficult for participants to manipulate prices. Markets are not explicit on expected returns, they do not weed out users who do not understand the decision rules or expected returns. Further, the participants have subjective probabilities with error. And, even though we saw no evidence of it in our study, markets can randomize end time to avoid the possibility of sniping at the last moment.[2]

Participants learning, over time and with higher incentives, how to maximize expected returns with manipulation is a potential issue in both laboratory and field. We are confident that we provide adequate stakes for a laboratory experiment. Prior work tested the validity of running experiments on Mechanical Turk and consistently found that even with stakes substantially lower than in traditional laboratory experiments, Mechanical Turk workers put reasonable effort in participating in economic experiments and produce results comparable to laboratory settings (e.g., Horton et al. 2011; Amir et al. 2012). For a further discussion on this topic, please see the Appendix in Electronic supplementary material. Our experiments paid an average hourly wage of around \$10 which is competitive in the Mechanical Turk labor market. In addition, we observed manipulation in some settings and none in others suggesting that behavior was guided by the experimental setting presented to participants, not by randomness. There was no increase in manipulation over eight rounds in the two-player version of the experiment; eight rounds with direct explicit feedback over 17 min is a serious amount of investments and feedback for the participants to not learn. Although, that does not exclude the possibility of learning and the value of future extensions of following participants over longer periods of time. Nor does it exclude the possibility of manipulation under great incentives, both within and outside of the markets, which can occur in a real-world setting.

There is no reason to assume that participants in real markets will be more expert at gaming the markets than the participants in this study. Most participants in company prediction markets are domain experts, not market experts. And, it takes a dedicated expert in markets to know how to manipulate. The authors of this paper, experts in how to manipulate a conditional decision market, were even tempted to not manipulate in one player deterministic treatments during the training runs for this experiment; it is just extremely engrained in most market investors to push the price towards their subjective probability. But, the more users and the more open the market, the more likely it would be subjected to some users with both market and domain expertise to induce manipulation. And, ultimately, the manipulation happens at the margin, not on the average.

[2] A good example of this is the American Civics Exchange which trades on political outcomes for real cash prizes. They are not conditional decision markets, but since they freeze their market once per month and provide prizes based on the performance of the participants with the current prices, they are subject to a very similar form of manipulation. In their first month they found out that traders were manipulating prices just under the deadline, which would only work if no other trader had time to counter and take the free money (in expectation). Thus, they switched to a random, unannounced, closing time on the announced final day: http://www.amciv.com/rules/.

In summary, contrary to theoretical work on conditional decision markets, our results suggest that manipulation may not occur, even when conditions favor it. These results add to the knowledge base of market design research; designers can draw on these insights and apply them to future corporate prediction markets for reshaping social decision making.

References

Amir O, Rand DG, Gal YK (2012) Economic games on the internet: the effect of $1 Stakes. PLoS ONE 7(2):e31461

Bennouri M, Gimpel H, Robert J (2011) Measuring the impact of information aggregation mechanisms: An experimental investigation. Journal of Economic Behavior & Organization 78(3):302–318

Berg JE, Nelson FD, Rietz TA (2008) Prediction market accuracy in the long run. International Journal of Forecasting 24(2):285–300

Berg JE, Rietz TA (2003) Prediction markets as decision support systems. Information Systems Frontiers 5(1):79–93

Bughin J, Chui M, Pollak L (2013) Organizing for change through social technologies: McKinsey Global Survey results. McKinsey & Company Inc, New York

Chen Y, Dimitrov S, Sami R, Reeves D, Pennock D, Hanson RD, Fortnow L, Gonen R (2010) Gaming prediction markets: Equilibrium strategies with a market maker. Algorithmica 58(4):930–969

Chen Y, Kash I, Ruberry M, Shnayder V (2011) Decision markets with good incentives. In: Internet and Network Economics, Springer, pp. 72-83

Chen Y, Kash IA (2011) Information Elicitation for Decision Making. In: Proceedings of the 10th International Conference on Autonomous Agents and Multiagent Systems (AAMAS'11)

Chinn S (2000) A simple method for converting an odds ratio to effect size for use in meta-analysis. Stat Med 19:3127–3131

Deck C, Lin S, Porter D (2013) Affecting policy by manipulating prediction markets: Experimental evidence. J Economic Behavior & Organization 85:48–62

Dimitrov S, Sami R (2008) Non-Myopic Strategies in Prediction Markets. In: Proceedings of the 9th ACM Conference on Electronic Commerce (EC'08), 200-209

Dimitrov S, Sami R (2010) Composition of markets with conflicting incentives. In: Proceedings of the 11th ACM Conference on Electronic Commerce (EC'11), 53-62

Gibbard A (1977) Manipulation of schemes that mix voting with chance. Econometrica 45(3):665–681

Gimpel H, Teschner F (2013a) Collective intelligence in market-based social decision making. In: Proceedings of the Thirty Fourth International Conference on Information Systems (ICIS'13)

Gimpel H, Teschner F (2013b) Prediction Markets with Experts Concerned in Post-Market Information Usage: Experimental Evidence on Collective Intelligence. Available at SSRN 2292116

Gneiting T, Raftery AE (2007) Strictly proper scoring rules, prediction, and estimation. Journal of the American Statistical Association 102(477):359–378

Hanson RD (1999) Decision markets. IEEE Intelligent Systems 14(3):16–99

Hanson RD (2003) Combinatorial information market design. Information Systems Frontiers 5(1):107–119

Hanson RD (2007) Logarithmic market scoring rules for modular combinatorial information aggregation. Journal of Prediction Markets 1(1):1–15

Hanson R, Oprea R, Porter D (2006) Information aggregation and manipulation in an experimental market. Journal of Economic Behavior and Organization 60(4):449–459

Healy PJ, Linardi S, Lowery JR, Ledyard JO (2010) Prediction Markets: Alternative Mechanisms for Complex Environments with Few Traders. Management Science 56(11):1977–1996

Horton JJ, Rand DG, Zeckhauser RJ (2011) The online laboratory: conducting experiments in a real labor market. Experimental Economics 14(3):399–425

Jian L, Sami R (2012) Aggregation and manipulation in prediction markets: Effects of trading mechanism and information distribution. Management Science 58(1):123–140

Ledyard J, Hanson R, Ishikada T (2009) An experimental test of combinatorial information markets. Journal of Economic Behavior and Organization 69(2):182–189

Mason W, Suri S (2012) Conducting behavioral research on Amazon's Mechanical Turk. Behavior Research Methods 44(1):1–23

McCarthy J (1956) Measures of the value of information. PNAS: Proceedings of the National Academy of Sciences of the United States of America 42(9):654–655

Oprea R, Porter D, Hibbert C, Hanson R, Tila D (2007) Can manipulators mislead market observers? University of California, Santa Cruz, mimeo

Othman A, Sandholm T (2010) Decision rules and decision markets. In: Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems (AAMAS'10), 625-632

Pennock DM, Sami R (2007) Computational aspects of prediction markets. In: Algorithmic Game Theory, Cambridge University Press, Ch. 26, 651-674

Rhode PW, Strumpf KS (2006) Manipulating political stock markets: A field experiment and a century of observational data. University of Arizona, Mimeo

Rothschild DM, Sethi R (2013) Trading Strategies and Market Microstructure: Evidence from a Prediction Market. Available at SSRN 2322420

Savage LJ (1971) Elicitation of personal probabilities and expectations. Journal of the American Statistical Association 66(336):783–801

Teschner F, Coblenz M, Weinhardt C (2011) Short-selling in prediction markets. The Journal of Prediction Markets 5(2):14–31

Wagman L, Conitzer V (2008) Optimal false-name-proof voting rules with costly voting. In: Proceedings of the National Conference on Artificial Intelligence (AAAI'08), 190-195

Wolfers J, Zitzewitz E (2006) Interpreting prediction market prices as probabilities (No. w12200). National Bureau of Economic Research