An Experimental Analysis of Information Aggregation in Decision Markets

A thesis presented in partial fulfilment of the requirements for the degree of

Doctor of Philosophy in

Economics

At Massey University,

Albany

New Zealand

Pansye ElKashef

2023

Abstract

Knowledge in a society is often distributed amongst different individuals, each holding different pieces of information. By aggregating these dispersed, different pieces of information, accurate forecasts can be generated, adding potential to improve decision-making processes. Decision markets are economic mechanisms to concurrently predict the future and decide on it. They incentivize "expert" individuals to predict the consequences of each of a set of possible actions and then select an action based on these predictions. Decision markets rely on scoring rules (payment schemes) to guarantee that experts are properly incentivized to truthfully reveal their beliefs whilst using decision rules to translate aggregated forecasts into decisions.

In this thesis, we present an experimental study of information aggregation in decision markets. Objective of the study is to provide a proof-of-principle for the functioning of decision markets. Market prices are dependent on the private signals given to participants, signifying that signal constellations are the primary determinant of final market prices. We find that decision markets work in aggregating private information and that the incentive compatibility of the decision rule matters for information aggregation. Upon exploring behavioural attributes that might be linked to individual trading performance, we discover that the decision rules also have an impact on participants' behaviour in the market but we find no evidence that individual behavioural attributes has any influence on market efficiency. Our findings can inform future experiments about decision making processes and real world decision market applications.

Acknowledgements

To Mum, this is because of you and for you.

To Dad, you would have been proud.

To Dr. Mohamed El-Komi, I discovered behavioral and experimental economics because of you. This is a result of your constant support and mentoring beyond the realm of academia.

As customary, I am indebted to my supervisors Thomas Pfeiffer, Ananish Chaudhuri and John Matthewson. Thomas's extensive mentoring, guidance and insights have taught me to challenge my own views. Thank you for giving me the opportunity to work in a highly interdisciplinary research environment that provided a stimulating environment to grow and learn. Ananish's encyclopedic knowledge as a behavioral and experimental economist has always left me in awe, I hope one day I would be as knowledgeable as you. I want to thank John for his excellent analytic abilities that taught me better communication skills. Particularly, I would like to express my gratitude to Addison Pan for her great coding skills that made this experiment possible, without you, I would have probably been lost. All were instrumental in shaping my own research in this area and none of them are responsible for any shortcomings of the final product.

The financial support of The Marsden Grant and Massey University is gratefully acknowledged. I wish to thank the many researchers I met for sharing their insights about laboratory experiments in market design. Specifically, I extend my gratitude to Simona Fabrizi, James Tremewan and Ronald Peeters for their many helpful comments and advice on the experimental design and their input on the early drafts relating to the body of the thesis.

Table of Contents

List of Figures

List of Tables

Defining the Scope of Work

"Did it ever occur to you that there might be more than one alternative?"-Richard Dormer in Good Vibrations, 2012

In today's society, knowledge is frequently spread, with various people possessing distinct pieces of information. If properly aggregated, this information can outperform that of any single individual, a concept known by the phrase of "wisdom of the crowds." There is considerable interest in developing mechanisms to harness this wisdom for forecasting and decision-making. These mechanisms include prediction markets.

Prediction markets are popular tools for aggregating dispersed information and producing highly accurate forecasts. Participants in prediction markets trade contracts with payoffs tied to the outcome of future events. The pricing of these contracts reflects aggregated information about the probabilities associated with the possible outcomes. Potential caveats of such an interpretation have been discussed in the literature but are not seen as critical for typical applications (Wolfers and Zitzewitz 2004).

Decision markets are mechanisms derived from prediction markets and work in a stepwise process to select one among several mutually exclusive actions. First, forecasts about the expected future consequences of each action are elicited in a step similar to incentivized prediction markets. Second, a decision rule is used to select an action based on the forecasted consequences. Once consequences of the selected action is revealed, payoffs are provided for the forecasts as elicited in the first step. Thus, while in a prediction market a set of participants make a prediction about an unconditional future outcome (e.g. whether a project gets completed successfully), decision markets forecast consequences of actions and select the action with the most favorable outcome (e.g. employ a candidate who is most likely to complete a project successfully). By making forecasts in the market, participants influence which action is going to be selected.

Motivation: Lack of Experimental Evidence

Decision markets have been developed in theory, but so far there have been very few applications. Significant empirical research is required to bring them into practice. Experiments appear to be most effective to inform practical design when used in conjunction with theoretical and empirical studies (Roth 2015). As such, what is currently needed most is a convincing experimental demonstration of the functioning of decision markets. This thesis provides such an experimental proof-of-concept.

To the best of our knowledge, and until recently, only two experimental studies have been published on different aspects in decision markets. The first experimental insights on decision markets were published in a PhD thesis by Forsell (2016) where the author aimed to evaluate whether the prediction accuracy of decision markets provided more accurate predictions than those of a prediction market using different combinations of scoring rules. The second experimental investigation of decision markets was by Teschner et al. (2017) where the authors conducted experiments on the Amazon Mechanical Turk platform to evaluate the impact of manipulation on decision quality. As discussed in detail in Chapter 2, these studies have generated empirical knowledge on decision markets but there is need for additional research.

Research Questions and Aims

Experiments can be designed to serve multiple objectives, such as demonstrations or to assess hypothesis. Depending on the experiment's objective, the design can also have a variety of aims. These variety of aims could be how much consideration is given to control the environment (e.g., ensuring that traders don't communicate their private estimations to other traders in the market) and how many settings are used investigating appropriate parameters to accommodate the experiment's objectives (e.g., number of traders, number of groups etc.). Either way, experiments have proven to be effective in informing practical design when used in conjunction with theoretical and empirical studies (Roth 2015). Our experiment primarily serves as a demonstration of a mechanism that is almost exclusively studied theoretically. It also investigates a number of testable hypotheses about the functioning of this mechanism to better understand the experimental demonstration.

Existing human subject studies on decision markets (Gimpel and Teschner 2014, Forsell 2016) have not provided a convincing proof-of-principle but rather point to potential problems with information aggregation. The aim of this thesis is to implement decision markets in a controlled laboratory setting with pre-established 'experts'. An analogue of Plott and Sunders's (1988) prediction market experiment will be used to provide an experimental demonstration of information aggregation in decision markets. Our market settings resemble potential real-world settings, where there is significant knowledge barriers and unexperienced traders are the norm. Our research is relevant in the domains of potentially improving evidence based and forecast based decision making in areas such as public policy development.

Our primary research question is: “Do decision markets aggregate information? If so, does their predictive performance match or outperform those of a prediction market’’. Exhibited trader characteristics and behaviors that may be associated with information aggregation will be investigated to help us understand if and how such characteristics may affect information aggregation. Furthermore, as a market needs proper incentives to succeed, hypotheses around providing incentives to traders are examined.

Overview and Structure of Thesis

The first part of my thesis introduces the key concepts from the literature behind information aggregation and the functioning of markets as tools for information aggregation. I focus on types of markets that have recently paved way to better-informed approaches of decision-making, such as prediction markets and more recently, decision markets. I chose to investigate the fundamentals behind the functioning of these specific markets and their performance because to understand how such markets are utilized as decision-making tools, it is vital to investigate how such markets originated, how they operate and how the incentives behind them work.

The second chapter of this thesis describes the design of an experiment to study how decision rules in decision markets affect information aggregation. I investigate whether the predictive performance of prediction markets can be matched by a decision market, and if so, whether it is sensitive to the decision rules used. To do so, I implement decision markets in a controlled laboratory setting with pre-established ‘experts’ that trade in markets employing different decision rules.

In the third chapter I describe the experimental results from the decision market experiment described in chapter two. I analyse how efficiently decision markets aggregate information, and whether their predictive performance matches that of a prediction market. I investigate my main research question by implementing a set of ordinary least-square (OLS) models to address two main hypotheses. The first hypothesis is whether final market prices depend on the private signals of participants and the second is whether the market errors depend on the decision rule used.

The fourth part of this thesis extends on the findings from the third chapter. Using data from the demographic questionnaire that is part of the experiment and results from the lottery task to assess risk preferences, I investigate if there are behavioral attributes associated with trading performance. I do so by addressing three hypotheses. The first hypothesis investigates if there is evidence of individual differences in the trading behavior of participants. The second hypothesis investigates if there are any individual differences in trading behavior that are connected to the individual attributes from the questionnaire and lottery task. The final hypothesis extends the findings of the second hypothesis by exploring whether any detected individual differences impact market accuracy. Main findings are summarized in the conclusion section of this thesis.

Chapter 1: Introduction - The Dilemma of the Hesitant Decision Maker

Forecasting and Information
Aggregation in Markets

"When my information changes, I alter my conclusions. What do you do, sir?"- Attributed to John Maynard Keynes by Paul Samuelson

1.1 Introduction

As outlined in the Scope of Work, this thesis is about an experimental study on decision markets. Decision markets employ market mechanisms for the purpose of aggregating information that is then used to make informed decisions. As such, this thesis involves aspects from information aggregation, markets, prediction markets and decision markets. The objective of this chapter is to introduce and provide a brief review of the relevant background of these concepts and terms to facilitate a better understanding of the experimental design and the results. The objective is not to systematically examine every paper on all relevant topics, but rather to highlight the key ideas and describe how relevant key concepts work.

In section 1.2, I provide a background on information aggregation as this is a key element for the functioning of prediction and decision markets. Then I provide a general introduction about the functioning of markets. A reason for doing so is that, to understand how such markets are being utilized as decision-making tools, it is vital to investigate how such markets operate and what incentives they generate. Only then, would we be able to grasp the importance and significance of these markets and be better equipped at sharing new research prospects. I focus on the types of markets that have recently paved way to better-informed approaches of decision-making such as, prediction markets and the more recent, decision markets. In section 1.3, I describe what markets are, their structure, design, and implementations. Section 1.4 discusses prediction markets and their relation to scoring rules and is followed by section 1.5 which explores decision markets. Finally, section 1.6 summarizes the key concepts of this chapter.

1.2 Information Aggregation

Everyday decision-making necessitates the integration of complex information. Decision-makers must integrate knowledge within the context of a present objective, comprehend the given circumstance, qualitatively assess numerous behavior alternatives, and pick the best option. Such a task is not restricted to individuals but could be expanded to groups as well.

It has been argued that amongst a group's many advantages, a group has greater problem-solving capacity than that of any one individual member (Kerr and Tindale 2004). In fact, most problems in our society cannot be solved by a single person but requires information from multiple individuals. The relevant knowledge is most likely dispersed across many individuals. If it is possible to aggregate and utilize this dispersed knowledge, one could enhance decision-making.

The first to use mathematical models to investigate information aggregation within a political sphere was Marquis de Condorcet (1785). The Condorcet jury model (CJM) was developed as a mathematical argument for the superiority of the majority rule in voting to aggregate dispersed knowledge amongst individuals who share common preferences. In his model, there are two states of the world, a and b. Each voter possesses privately known information, in the form of a binary signal, α or β, as to which state of the world is most likely. The voters must collectively decide on one of two possible policies, A and B, where all voters prefer A in state a and B in state b. If the individual signals prove to be informative (e.g., $p\,\{\alpha|a\} = p\,\{\,\beta|b\} = q > 0.5$), then as the number of voters increase, the likelihood that the majority vote would reach the correct decision would be close to 1. [1] The CJM demonstrated that even though the true state of the world is unknown to any voter, each individual holds a piece of private information about the underlying state and by accumulating this dispersed private information, more accurate decisions would be achieved. This model has endured for more than two centuries and remains the workhouse model of information aggregation through voting mechanisms.

Two popular approaches that have studied and addressed different ways of aggregating information are, wisdom of the crowds and collective decision-making. The concept of ''Wisdom of the Crowds'' was examined by Surowiecki (2004). In his book, Surowiecki reviews numerous examples to demonstrate that aggregate group estimates are collectively smarter than that of any single individual and that the best forecasts come from crowds. A classic well-known example of this is a 1907 experiment by Francis Galton in which he questioned 787 locals about the weight of one ox. Each villager reported a different and incorrect response. Galton, on the other hand, averaged these estimations and accurately determined the weight of the ox. This popular example demonstrated that people in crowds can pool their information to demonstrate collective intelligence.

Collective decision-making is the process through which group members adjoin and agree on a plan of action (Montes de Oca et al. 2011). Also known as a ''consensus decision'', it takes place when all the members of a group collectively select the same action from a set of mutually exclusive alternative actions (Conradt and Roper 2005).

[1] Where all voters vote for the policy they personally think is best given their private information.

An example of collective decision-making could be seen amongst the members of United Nations Climate Change Conference when they gathered in 2015 to agree on a target of how to limit global warming to "well below 2°C" (Boehmer-Christiansen 2016). This conference revealed that collective actions based on aggregated information is powerful and can outperform actions based on information possessed by a single individual.

The above two concepts of information aggregation have been frequently studied individually, especially to develop information aggregation algorithms. However, recent research has started to emerge by comparing and integrating both methods to try to achieve the optimum aggregation performance. An experiment by Lee and Shi (2010) was conducted to investigate people's cognitive abilities to predict the price of household goods. Under different experimental settings, individuals were asked to estimate the price of common household items which they are familiar with but were unlikely to possess the accurate price information. The aim of the task was to evaluate the performance of several small group estimates in comparison with the traditional Wisdom of Crowds analysis. Results of the experiment showed that that simple aggregation through averaging individual estimates in line with the Wisdom of Crowds approach outperformed all other types of estimation settings.

On the other hand, in a recent study by Navajas et al. (2018), the authors asked a large crowd of people that were attending an event a series of general knowledge questions (e.g., "What is the height of the Eiffel Tower?'') and the crowd were asked to provide answers to those questions as estimated quantities. People were first asked to answer the questions individually. Later, they were divided into groups of five to make consensus decisions about the questions through revising their own individual estimates. The study found that averaging the consensus estimates of the questions from the groups provided more accurate estimates than aggregating the initial individual estimates. In other words, a process that involved elements of collective decision-making outperformed a simple aggregation of initial estimates.

The findings of the above two studies indicated that by incorporating both approaches of information aggregation, more efficient and accurate strategies for extracting desired outcomes could be achieved. However, deciding whether to use one or two information aggregation methods does not undermine the significance of either technique. In fact, regardless of which technique is used, information aggregation amongst a group of individuals could surpass that of a single individual.

1.3 Market Structure, Design, and Implementation

According to Ellis (1990), the term 'market,' as used by economists, has a slightly different connotation than it does in everyday use. It does not refer to the actual location where products or services are sold or bought (just like in a local "village market"), nor does it refer to the phases that a product goes through between the producer and the consumer (as in production of marketing channels). Instead, in an abstract sense, it describes the buying and selling of a commodity and the development of its price. To narrow things somewhat, in this context, the phrase encompasses the numerous decisions made by commodity producers (the supply side of the market) and commodity consumers (the demand side of the market), which together determine the commodity's price level (Hayek 1945).

1.3.1 Structures

Early discussions by Hayek (1945), subsequent investigations by Muth (1961) and developments by Grossman (1976), have been the central foundation for how economic thought is constructed on the concept that markets can aggregate dispersed private information into prices. This idea is also the central theme of the ''efficient market hypothesis'' (Fama 1970). More details are provided in section 1.4.1. Market prices, according to Hayek, served the aim of disclosing common and private knowledge, whereas Muth demonstrated that a market price incorporates all the available information from the participants.

The operation of markets is characterized by a variety of structures. The continuous double auction (CDA) is the most conventional for financial markets where buyers place "bid" orders and sellers place "ask" orders at a specified price and volume (Das et al. 2001). When there is a price match, or when the asking price is lower than the bidding price, a trade occurs. The transactions proceed until all possible orders have been satisfied and a spread occurs between the prices of asks and bids. During the trading phase, information is constantly updated, and traders can make a profit by using news and events to better judge the value of the traded assets. However, if there exists a large spread between the bid and ask prices (due to only few traders being in the market), trading might not occur. In such cases, some markets use a market maker to reduce such a spread and to ensure sufficient liquidity.

A market maker (MM), also known as a liquidity supplier, is essentially a market participant who quotes both a purchase and a sell price for a tradable asset to profit from the bid–ask spread. The MM ensures that there are always some assets in the market for which a price is made available. As such, the MM's role is to assist in minimizing price variation (volatility) by establishing a defined trading price range for certain assets. In some markets, the MM plays a dominant role, with traders typically (or primarily) interacting with the MM rather than another trader. This is the case, for instance, with pari-mutuel markets.

Pari-mutuel markets have originally been created for sports betting but are now used for the forecasting of future events beyond this domain (Plott et al. 2003). In a pari-mutuel market, traders acquire assets from the MM at a fixed price and all money "bets" for individual assets are pooled and subsequently awarded to the final winners. The ratio of the money bet on one specific asset to the total pool of bets determines the payoffs of a bet and can be interpreted as the perceived probability for an asset to win. Pari-mutuel markets are not constraint by liquidity; thus traders are guaranteed that their orders will be taken. Furthermore, in contrast to other market mechanisms that are used for sports betting (such as bookkeepers), in a pari-mutuel market, the MM carries no risk and is safeguarded from incurring a loss. However, in such markets, no method exits in which a trader can sell or bet back at a profit (assuming that the bet was on an asset whose true probability increased); a trader who receives information or learns from the market cannot receive an advantage. Hence, all rational traders should wait and trade only just prior to the closing of the market to have complete information. This implies that the market will not necessarily reflect complete information until shortly before the market closes (Fama 1965).

Further mechanisms have been established that attempt to combine the best measures of pari-mutuel and CDA markets. For example, the Dynamic Pari-Mutuel market (DPM) has been created with the aim of being a genuine hybrid of both the CDA and pari-mutuel market by incentivizing experts to reveal their information early (Pennock 2004). Moreover, Robin Hanson (2003) has proposed markets utilizing market scoring rules (MSR). These MSR's can be used by market makers in prediction markets to forecast a single variable such as the likelihood of an event occurring (Christiansen, 2007). The relation between MSR's to traditional scoring rules will be described in section 1.4.3.

1.3.2 Early experimental studies on markets

Market design is an economic tool that has grown to encompass the design of marketplaces, economic settings, institutes, and allocation rules (Kagel and Roth 1995). It does not merely incorporate the creation of new institutions, but also highlights how the architecture of economic institutions affects their performance. It is both one of the oldest and one of the most recent branches of experimental economics. It is claimed to be one of the oldest as every economic experiment encompasses the design of an economic environment, with numerous experiments comparing the outcomes and effects of various designs. It is also one of the most current branches due to its re-emergence in the 1990's, particularly in dealing with novel challenges such as matching residents to medical colleges or kidney recipients with kidney donors without monetizing the assets in the market (Roth 2015).[2]

Chamberlain (1948) developed the first market experiment in economic history, demonstrating how market design continuously plays a role in experimental economics. He explored competitive equilibrium by designing not just the type of marketplace - pairwise negotiation - but also inventing an extensively used strategy that until today, remains the main inducement for supply and demand curves. He wanted to illustrate how deviations from perfectly competitive equilibrium could arise.

[2] The challenges for those markets emerged as a result of the importance of exchanging goods for money.

This strategy involved endowing sellers and buyers with prices and quantities to effectively sell or buy units of goods from the experimenter to complete any trade transaction. For example, in an experimental market, one subject was informed that he could sell one unit of some good in the market that would cost him $20 (which will be deducted from his sale price) whilst another subject was told that he would be paid $50 for the first unit he bought (minus his purchase price). If those two subjects transact at a price of $30, the seller would earn $10 while the buyer would earn $20. As such, Chamberlain used imputed values of the willingness to pay on the part of buyers and the willingness to accept on the part of sellers in creating an experimental methodology de novo, that added significant empirical power not only to microeconomics, but to market design.

Meanwhile, the notion of Nash equilibrium proposed by John Nash (1950) has proved to be a good behavioral predictor, both in experimental and field settings (Maskin 2011). Nash showed that the way an experiment is designed, for example, whether the experiment had repeated or non-repeated interactions, would impact participants behavior. An example for the relevance of the Nash equilibrium for market design is the Vickrey auction or sealed-bid second-price auction (SBSPA). A Vickrey auction is a type of sealed-bid auction where bidders submit sealed one-time offers, and the highest bidder wins but at the price offered by the second highest bidder. This type of auction is important for market behavior as it introduces incentive compatibility in a market. The SBSPA mechanism is incentive compatible in a sense that if participants act according to their true preferences (honest bidding), they will achieve the best outcome. As such, resources are allocated efficiently, which is a desirable characteristic of a market.

1.3.3 Practical implementations

Airline deregulation paved way to one of earliest approaches of utilizing experiments for practical market design. Following the establishment of the Airline Deregulation Act in 1978, the aviation sector underwent tremendous transformation. Many areas of the airline industry that were formerly controlled such as ticket pricing, which airlines flew to which locations, and how new airlines entered the market have been deregulated, allowing individual airlines to make decisions and compete with one another.

In 1979, the Civil Aviation Board (CAB) appointed experimentalists David Grether, Mark Isaac and Charles Plott to assess the effectiveness and efficiency of slot committees. They were asked to solve the problem of the allocation of take-off and landing slots at crowded airports - an issue that affected millions of travelers each year (Grether et al. 1981). Grether et al. were the first to suggest a market-based solution to the slot congestion dilemma. Their research unveiled to the aviation board that there are market-based approaches to allocating slots that are generally more efficient than committee-based allocations.[3] The way the experimentalists used economic experiments as a viable source for obtaining meaningful, controlled data was a unique element of applied economics research at the time.

Another example of the important role that experiments played in practical market design was in the early 1990's. In 1994, the Federal Communications Commission (FCC) organized several electromagnetic spectrum auctions (Plott 1997). These auctions were simultaneous ascending auctions (SAA) in which sets of licenses were auctioned

[3] At that time, slot committees used unanimity voting rules (Grether et al., 1989).

over several rounds at the same time.[4] They were publicly accessible to any qualifying individual or firm that could submit an application accompanied by an upfront deposit.[5] New Zealand was one of the many countries that utilised SAA auctions to competitively allocate electromagnetic spectrums and has entirely adopted such auctions to efficiently allocate spectrums to those who valued it the highest (Review of Radio Spectrum Policy in New Zealand 2005). Soon after, the FCC considered adopting a combinatorial auction to replace those simultaneous, round based auctions (Plott and Salmon 2004). The experimental results of adopting these auctions indicated that the continuous phase was responsible for most of the adjustment and efficiency, which prompted a re-examination of combinatorial auctions.[6]

An experiment can perform different roles, such as designing or testing, to bring forth new market conceptions to its implementation phase. That was the case with an experiment conducted in the Fall of 2011 at the University of Pennsylvania's Wharton School by Eric Budish and Judd Kessler. The school's committee convened to re-evaluate its ''fake-money'' auction mechanism for assigning MBA course schedules (Sönmez and Ünver 2003, 2010). Such auctions were widely used at many educational institutions.[7] It was discovered that the mechanism had significant inadequacies, resulting in inefficient and unjust allocations (Budish et al. 2015). To address these problems, it was recommended that a new course-allocation method, known as the ''Budish (2011)" mechanism to be employed, simultaneously accompanied by a planned laboratory experiment to verify the mechanism's appropriateness for practice.

[4] They were also known as simultaneous multiple round (SMR) auctions or the simultaneous multiple round ascending auction.
[5] The FCC determined the qualified bidders.
[6] Previously, the Commission relied heavily on comparison hearings and lotteries to choose one licensee from a pool of mutually exclusive candidates for a licence.
[7] See Sönmez and Ünver (2010) for a list of schools that adopted such auction mechanisms.

Later, an improvement to the Budish (2011) system was created and deployed, now known as Course Match, which improved several aspects of the prior market structure. The experiment effectively transitioned the new market design from theory to practice. These difficulties prompted the development of a unique experimental design that used real market participants, evaluated their ability to properly report complicated preferences, and looked for unexpected effects that the theory could have overlooked. The new mechanism outperformed the old one, demonstrating that when a previously unknown innovative design is suggested, experiments can generate useful empirical data that is not available elsewhere.

Overall, the examples provided above indicate that in the past, an interplay between experiments and theory have been very successful in developing mechanisms that are used in practice for highly relevant real-world problems. For decision markets, my research is positioned relatively early in such a development from theory to implementation. Although theory exists, these mechanisms have received little experimental investigation.

1.4 Prediction Markets

Prediction markets (also called information markets) provide a way of aggregating private beliefs about future events. Most markets are related to contemporaneous transactions and their price discovery, but prediction markets involve transactions and price discovery of an event that will transpire only in the future. They are a place where information is aggregated via market mechanisms for the primary purpose of forecasting events, e.g., the probability that an event will occur. They are a viable example of employing the “wisdom of the crowds” where one can enhance decision-making by aggregating and using dispersed knowledge (Surowiecki 2010).

The notion that prediction markets effectively aggregate dispersed information about future events has long been recognized and is based on the ‘’efficient market hypothesis’’ (Chen and Penncok 2010, Wolfers and Zitzewitz 2004).[8] The efficient market hypothesis asserts that aggregation is ‘’perfect’’ in a sense that financial markets are informationally efficient, and that agents’ expectations are equal to the market prices which ‘’fully reflect all available information’’ about future outcomes (Berg et al. 2008, Bothos et al. 2012, Gruca et al. 2005, Hanson et al. 2006). The special features of prediction markets set them apart from standard forecasting approaches and thus, aid in guiding decisions.

Market structures are employed for their informative and predictive purposes, and “prediction markets” is one of several terms that are used to characterize them. Since the idea of such markets were first introduced in the early nineties (Hanson 1990) there has been extensive theoretical work on how to specify assets to aggregate traders’ predictions (Hanson 2003, Iyer et al. 2010, Chen 2011, Ostrovsky 2012); how to interpret the market results (Manski 2006, Wolfers and Zitzewitz 2006); and how to use market scoring rules to incentivize traders to reveal their information to the market (Hanson 2007, Chen et al. 2012).

Instead of asking individuals to each reveal their beliefs separately (using either incentivized or non-incentivized mechanisms) and conjoining them according to some rule, in a market, this same belief elicitation can be achieved by asking individuals to

[8] The article by Fama (1970) gives substantial empirical evidence for the efficient market hypothesis.

engage in trading assets with payoffs tied to the outcome of these future events. The participants then have an incentive to truthfully reveal their beliefs by weighing their personal opinions through the amount of money they invest into an asset. As a result, when forecasting future outcomes, traders ''put their money where their mouth is" (Hanson 2009).

Prices of the traded contracts reflect the aggregated predictions of traders about the probabilities associated with the possible outcomes (Plott and Sunder 1988). To properly elicit the necessary information in a prediction market, determining the correct way to design traded contracts is necessary. One of the methods to structure such contracts would be the popular “winner-take-all’’ model, in which two or more contracts are listed in the market. The market price of the contract will represent its chances of occurring and if it does, only that contract would be paid out. Arrow-Debreu securities are an example of those frequently used contracts. They pay $1 if a specific outcome is realised and $0 otherwise.

Other contract types, such as “forward” contracts and “spread” betting contracts, can also be used to predict outcomes of future events (Wolfers and Zitzewitz 2004). In a forward market, contracts are paid out on the exact state of a specific value at a specific time. Such forms of futures markets are commonly found with standard exchanges, such as the cost of an oil barrel being delivered in 60 days. In spread betting, traders differentiate themselves by bidding on the cutoff that determines whether an event happens (Wolfers and Zitzewitz 2004). For example, in basketball, point-spreading is a type of spread betting that could be used in which the bet is either that one team will win by at least a certain number of points, or not. Under this type of betting, the price of the bet is fixed, but the size of the spread is flexible.

The early Iowa Electronic Market (IEM) was developed in 1988 by The University of Iowa. Simple markets were established for the purpose of predicting the winners of political elections. Such markets were unique by creating a trading environment where traders could purchase contracts to predict whether George H.W. Bush would win the 1988 presidential election or not. Traders would purchase contracts through a web-based platform that allowed them to convey their predictions with a reward if their predictions held true.[9] The prices of these contracts would therefore be interpreted as a market probability of George H.W. Bush winning the re-election. For example, if the "Bush contract" was traded at $0.70 per contract, this expressed that the market evaluated Bush winning the re-election at a 70% chance. Despite the lack of profit incentives, since trading was restricted to a minimum financial amount because of U.S. securities regulations, the results from the IEM exhibited that future contracts for speculative, non-financial events are significant. Markets remained efficient and contract pricing predicting the probability of future events were accurate (Manski 2004).

1.4.1 Prediction markets with automated market makers

It can be challenging to coordinate transactions of predictive contracts across numerous traders and make sense of what the trades are saying about the future. Aggregating trade data from various transactions to get a clear assessment of the likelihood of an outcome becomes increasingly challenging as well. These are difficulties that can be mitigated by using an automated market maker (AMM).

[9] Furthermore, The Iowa Markets provided linear vote-share contracts that pay $1 times a candidate's vote percentage. As such, researchers have interpreted the price of a given contract as a ''best point" market-generated estimate of the candidate's vote share.

Traders can purchase and sell result tokens in a prediction market if the AMM is operational, and the AMM will aggregate the trade data to create estimates for the probability of outcomes. The logarithmic market scoring rule (LMSR) market maker, originally proposed by Hanson (2002), is one of the most common and widely used market makers in prediction markets. When traders buy/sell a certain quantity of contracts from an LMSR AMM, they convey the cost amount to the MM, and the MM uses a LMSR price function within the market mechanism to calculate the probability of future events occurring (by transforming the cost amount to a set of outcome contracts).[10]

1.4.2 The relation between scoring rules and market scoring rules

Market scoring rules are closely related to scoring rules, which have been used to measure the accuracy of forecasts and allow to incentivize single forecasters to provide accurate forecasts. Over recent years, the concept of single-expert scoring rules have been expanded to scoring rules that are sequentially used to elicit forecasts from multiple experts (Hanson 2003; Pennock and Sami 2007). Assume we have n mutually exclusive outcomes (1,…,n) of a future event. A forecast $\vec{r}$, would be a vector with n probabilities that add up to 1. A scoring rule is a function that maps the forecast $\vec{r}$ and the observed outcome i onto a real number S $(\vec{r}, i)$:

$$S\,(\vec{r}, i) \mapsto \mathbb{R}$$

A proper scoring rule is a scoring rule which can be used to reward an expert for her forecast. The reward is designed such that a risk-neutral expert maximises her

[10] Another popular price function used by AMMs are proper scoring rules such as the Brier scoring rule (1950). See section 1.6.

expected payoff by honestly reporting her beliefs. In other words, proper scoring rules provide an incentive compatible mechanism to elicit an expert's belief (McCarthy 1956; Savage 1971). They also provide incentives to experts to obtain information they would not possess otherwise (Clemen 2002). Proper scoring rules have long been used in economics forecasting (O'Carroll 1977), economics experiments and risk analysis (DeWispelare et al. 1995).

The two most common proper scoring rules are the Brier scoring rule introduced by Glenn W. Brier (1950) followed shortly by the Logarithmic scoring rule (Good 1952). The Brier scoring rule is given by:

$$S(\vec{r}, i) = a_i + b\left(2\vec{r}_i - \sum_j \vec{r}_j^{\,2}\right)$$

and the logarithmic scoring rule is given by:

$$S(\vec{r}, i) = a_i + b\log(\vec{r}_i)$$

Here j is an index running from (1,…,n); a_i are constants that move the score either up or down, depending on the outcome, and b > 0 is a constant multiplier that allows to scale the payoff. The a_i can be selected large enough such that experts are incentivized to participate voluntarily, i.e., the score meets the rational participation constraint. Such a scoring rule ensures that experts are better off by participating rather than not participating.

The logarithmic scoring rule implies that an expert's payoff depends only on the probability that she assigned to the actual realized outcome (Winkler 1969). It is the only proper scoring rule with this property. In contrast, the Brier scoring rule, looks at the entire report of the expert. Both scoring rules are strictly proper. This means that the expert's score is uniquely optimized by reporting the true probabilities.

Simple scoring rules are used to elicit experts' probability estimates. An expert reports a probability for each event and gets paid depending on that report and the actual event (Hanson 2003). As such, the expert interacts once with the scoring rule, and by reporting her honest beliefs, she will maximize her expected score.

To expand this to multiple experts, each expert could be scored according to their prediction with a fixed proper scoring rule. This, however, is costly because several experts would be paid for the same prediction and for an infinite number of experts, the costs would be infinite. To overcome this drawback, experts could alternatively be asked to report their predictions sequentially. After each report, the scoring rule changes depending on that report. This can be done by altering the constants a_i in the scoring rule accordingly. With such dynamic scoring rules, experts can be incentivized to improve on the predictions of previous experts. This solves the above drawback by having finite costs even for an infinite number of experts (Hanson 2003).

Sequentially updated scoring rules can pave the way into market scoring rules by incorporating the advantages of scoring rules and prediction markets. Formally, a market scoring rule consistently has a current probability distribution $\vec{p}$ that's equivalent to the last report any expert has made to it (Hanson 2003). Any expert can create a new report by altering a part of this distribution at any given time. An expert is obliged to increase some probability values and decrease others to amend a market scoring rule distribution $\vec{p}$. If she increased the probability and the true state turned out in her favor, she will gain, whilst if the true state turned out one where she has decreased the probability, she will end up losing. Thus, she is essentially making a bet with the market scoring rule about the true state, and since she can choose to alter the distribution $\vec{p}$ by merely a small amount, these small bets are all “fair bets” at probabilities $\vec{p}$ (Hanson 2007). This in effect lets her make any infinitesimal fair bets at the odds in the last report with no need to find another expert willing to make a matching bet. If the bets are implemented by trading Arrow-Debreu securities, and a logarithmic scoring rule is used, a market scoring rule will price assets in outcome i according to:

$$p_i(S) = k.\frac{e^{(a_i+S_i)/b}}{\sum_{j=1\ldots n} e^{(a_j+S_j)/b}}$$

Where p_i represents the asset price for an event i; S_i is the total number of assets sold for event i; k is the value of an asset in i if event i occurs; a_i is a constant adjusting the asset’s initial price; and b is a parameter that adjusts how fast prices move in response to asset purchases. As b measures liquidity and the potential loss of the market maker, a large b value allows a trader to purchase numerous assets at the current price without drastically affecting the price itself. The above equation assures that the scores are bounded between 0 and 1 and always sum to 1. Trading can be done with an automated market maker who facilitates trading. The above market scoring rule applies only to infinitely small trades. To trade a finite amount of assets, the expression has to be integrated.

In prediction markets, participants can make multiple trades and select the timing of their trades. This in principle allows for strategic manipulation. By deferring her truthful beliefs, an expert can manipulate her fellow traders within the market (Dimitrov et al. 2008; Chen et al. 2010). This problem has been studied and has shown that if participants have independent pieces of information, elicitation is myopically incentive compatible. Page (2011) discusses empirical and theoretical evidence about how elicitation of conditional probabilities is prone to manipulation. He points out that since the value of conditional probability estimates are vital to decision-makers, traders could be more motivated to manipulate these markets to influence the decision-maker's choices.

1.4.3 Practical applications

The performance of prediction markets has been examined in a variety of contexts and are used in a variety of settings. They were first suggested in the science domain in the early 1990's by Hanson and were tested in the laboratory shortly after by other researchers. Important contributions to real-world applications of prediction markets came from John Ledyard, starting with his research on forecasting the economic and political stability of the Middle East (Polk et al. 2003). The intriguing properties of these markets have driven him and other co-authors to examine their capacity to aggregate information in complex real-world settings (Ledyard et al. 2009) whilst utilizing laboratory experiments to capture information and intentions of individual traders that cannot be typically observed in field data (Bossaerts et al. 2013).

Numerous businesses have been experimenting with internal prediction markets to evaluate forecasting accuracy of key events to their firms. Best Buy, Chrysler, General Electric, InterContinental Hotels, Nokia and Pfizer are among the companies whose internal prediction markets have been mentioned in the public domain (Cowgill et al. 2009). Other well-known firms such as BP and Siemens Germany have used prediction markets for resource allocation and project management, respectively (BP Climate Change Manager Jeff Morgheim 2000, Ortner 1998). The outcomes of these markets are rarely disclosed since they frequently include sensitive information (Graefe and Weinhardt 2008).

Other corporations and individuals have utilized prediction markets for different reasons. Dreber et al. (2015) used prediction markets to forecast replications for a subset of The Reproducibility Project: Psychology (RRP) research (Open Science Collaboration 2015).[11] Traders were asked to bet on the binary outcomes of replicating 44 studies before the replication outcome was known. Their results showed that prediction markets accurately forecasted replication results whilst demonstrating that such markets are a viable method for determining the reproducibility of published scientific results at a low cost.

Prediction markets have also been popular amongst predicting sporting competitions. Spann and Skiera (2009) compared the accuracy of prediction markets to the performance of single experts and betting odds in forecasting the outcomes of German Premier League soccer games. After examining data from 678 games, the authors discovered two important findings. The first was that prediction markets performed as well as betting odds. The second finding was that prediction markets were more accurate than the predictions of single experts.

Further researchers have also tested the accuracy of prediction markets for sporting events. In an empirical study by Luckner et al. (2009), the authors compared the prediction accuracy of a prediction market for the FIFA World Cup 2006 to forecasts acquired from the FIFA World Rankings and a random predictor model. Their results showed that the FIFA World Cup prediction markets outperformed both scales. In

[11] Replication here refers to ''re-analysis replication'' in which reported results in a paper or manuscript can be verified by reproducing the same results using the same data. See Open Science Collaboration 2015 for more discussions about the different forms of replications.

another study by Wunderlich and Memmert (2016), the authors investigated the capability of a novel framework to derive more precise model-based predictions from the FIFA ranking systems. It included data derived from the World Cups of the years 2006, 2010 and 2014. These were compared to the forecasting performance of betting markets. It was revealed that despite both predictors being beneficial, the prediction abilities of the FIFA World Ranking have significantly improved and considerably outperformed the betting market for the 2014 World Cup.

All the above practical applications shed light on the potential of prediction markets to pave the way to better-informed approaches of decision-making. Their excellent forecasting performance have been the motivation behind numerous decision makers employing them to inform decision-making.

1.5 Decision Markets

An important use of prediction markets has been their substantial capacity to inform decision-making. By eliciting predictions about future events, prediction markets can be used as tools to guide decisions. However, there are instances when the probability of realized outcomes are dependent on a decision maker's action. In such a situation, a decision maker requires conditional forecasts, i.e., what will happen if a particular action is taken. Such conditional forecasts are characteristic for decision markets. Decision markets elicit forecasts that are *conditional* upon the executed decision (Chen et al. 2011, Teshner et al. 2017), and use these forecasts to inform decisions. In other words, they simultaneously “predict and decide’’ the future.

A descriptive example which showcases the use of decision markets comes from Chen et al. (2011); consider a project manager deciding on whether to hire developer A or B for her team. The main goal is for the project to be completed on time, which may be unequally likely for both candidates. Suppose the manager has access to knowledgeable experts, e.g., former employers of the two developers. To get a more informed decision, she requests both experts to take part in two conditional prediction markets: one forecasting the probability to finish the project on time should developer A be hired, the other should B be hired. Let's assume developer B is hired, the manager can then observe the project's development and reward the experts according to their level of accuracy in predicting the likelihood given B. A decision market thus requires an individual to observe what they ask experts to predict, act by hiring only one developer, and reward experts on the outcome (whether the project was completed on time).

While experts are employed to predict and aggregate information about future outcomes; the final decision usually remains with the decision maker, e.g., the project manager. The decision maker determines which decisions should be executed depending on the highest probability of an outcome occurring (accumulated from the experts). This is usually represented through market prices. Following through the example above, the project manager invited the experts to participate in two conditional prediction markets to convey their predictions about which developer to hire (A or B). Experts trade assets within these markets to represent their beliefs about which developer would complete the project on time. Now, let's assume that the final market prices for developer A is \$0.4 and \$0.7 for B. This signifies that the experts' likelihood beliefs of the project being completed by A is 40% and by B 70%. In this case, the decision maker can make an informed decision about which developer to hire. This leaves open the question of how to incentivize experts to report truthful forecasts.

Whilst generating markets to aggregate forecasts about future outcomes may seem simple, providing incentives to experts to truthfully reveal information has proven to be problematic (Othman & Sandholm 2010). It was initially contemplated that a decision maker could specify decision rules that link market forecasts to the available actions (Chen et al. 2014). Such decision rules would foster an environment for traders to confidently and truthfully trade according to their beliefs whilst allowing the decision maker to select the most profitable outcome, or the highest expected utility, out of all available outcomes.

An example of a decision rule is the deterministic decision rule. A deterministic decision rule implements the most profitable action given the forecasts. As some outcomes cannot be observed due to non-executed decisions, one way to address these non-executed decisions would be to void the corresponding assets and provide no financial reward for the expert (Othman and Sandholm 2010, Chen et al. 2011).[12]

In such a setting, providing proper incentives to reveal predictions on an outcome can become problematic. Othman and Sandholm (2010) show that using strictly proper market scoring rules (as implemented for market-making in prediction markets) are not sufficient to guarantee that traders have incentives to trade according to their beliefs in decision markets. Moreover, they show that deterministic decision rules are incompatible with providing proper incentives to elicit predictions. To address this issue, Chen et al. (2014) introduced the concept of strictly proper pairs which combine strictly proper scoring rules with stochastic decision rules. Chen et al. (2014) demonstrated that markets implementing such a strictly proper pair recover the

[12] As a decision market requires you to observe what you ask people to predict (they are conditional markets), non-executed decisions would be void. For example, if the decision maker chose to hire developer A and not B and one expert voted for A when B was selected, he would not be paid for all the assets owned on A.

desirable properties of prediction markets, where the best strategy of a trader is to invest exactly according to their beliefs.

Such a decision market brings together market scoring rules and stochastic decision rules. A stochastic decision rule reviews the market and assigns a probability to each action. A market scoring rule is used to define the experts' payoffs. If Arrow Debreu securities are used in the market, the assets on unobserved outcomes are void. Assets on the observed outcomes are scaled by 1/p, where p is the probability of the selected action. For example, if an outcome is selected with 25% chance, the payoffs will be increased by a factor of 4. If an outcome is selected with 75% chance, the payoffs will be increased by a factor of 1.33 (4/3). Intuitively, experts will be compensated for the risk of an unobserved outcome that could have possibly been observed and paid out. The expected payoffs of such a decision market are equivalent to the expected payoffs of a prediction market, and therefore findings on the incentive compatibilities of myopic traders in prediction markets carry over (Chen et al. 2011).

Theory suggests that decision markets can be strategically and mechanistically complex. Under a deterministic decision rule, a principal's decision can be deliberately manipulated by market participants. A stochastic decision rule can void this incentive; consecutively resulting in manipulation-free forecasts. As witnessed in voting mechanisms, randomness can lead to incentive-compatibility (Gibbard 1977; Wagman et al. 2008). This notion has been applied to a single expert with conditional scoring rules, demonstrating that incentive compatibility can be restored through a probabilistic decision rule (Chen et al. 2011). Nonetheless, this comes at a cost of the principal willingly and knowingly taking a sub-optimal decision with positive probability.

1.6 Conclusion

Information aggregated from a group of individuals outperforms information gathered from a single individual; especially when predicting outcomes. Whether it is estimating market growth or discovering the optimum strategy to beat a competitor, evidence shows that combining numerous, independent opinions is frequently more accurate than an expert's individual opinion. Numerous economists were frequently involved in the detailed design of markets and other economic organizations in ways that have transformed original ideas to the implementation of new designs, and even paving the way of new scientific literature to the design of markets. The increasing applicability of market design had resulted in the emergence of new experimental applications.

The fundamental concept of prediction markets is to allow individuals to trade their expectations about the outcome of a future event in a virtual market (e.g., predicting the number of flu vaccines being rolled out for 2021). They have proven to be remarkably potent as information aggregation mechanisms and frequently outperformed other information aggregation mechanisms, such as opinion polls and surveys. The ideas and use of prediction markets have proven to not be restricted to the academic domain, as they have been successfully used to forecast events in practical applications such as political events, climate, sports and even entertainment predictions.

The developments of prediction markets to guide decisions have paved the way to enhance their use to create more sophisticated markets, decision markets. The significant difference between decision markets and prediction markets is that they "predict to decide the future" (Chen et al. 2011). Predictions in decision markets must be conditional on the decision being carried out, hence termed "conditional decision markets" (Hanson 1999, Berg and Reitz 2003). It requires observing what experts are asked to predict and selecting only one outcome.

In this chapter, we have concentrated on the types of markets that have paved the way for more informed decision-making techniques, such as prediction markets and decision markets. Exploring the fundamentals behind the operation and success of these specific markets help us understand how such markets are used as decision-making tools. Understanding how these markets work, and the incentives that drive them, allows to identify gaps in the current knowledge on decision markets and design promising approaches to study these mechanisms.

References

Ackert L.F., Church B.K., Shehata M. (1997) Market behavior in the presence of costly, imperfect information: experimental evidence. J. Economic Behavior. Organ 33, 61–74

Adams C. (2006) Learning in Prediction Markets. Available at SSRN: https://ssrn.com/abstract=923155 or http://dx.doi.org/10.2139/ssrn.923155

Allais M. (1953) Le comportement de l'homme rationnel devant le risque, critique des postulats et axioms de l'ecole Americaine. Econometrica 21: 503–46

Arrow K.J., Forsythe R., Gorham M., Hahn R., Hanson R., Ledyard J.O., Levmore S, Litan R, Milgrom P, Nelson FD, Neumann GR, Ottaviani M, Schelling TC, Shiller RJ, Smith VL, Snowberg E, Sunstein CR, Tetlock PC, Tetlock PE, Varian HR, Wolfers J, Zitzewitz E (2008) Economics. The promise of prediction markets. Science 320(5878):877–878

Berg J.E., Nelson F.D., Rietz T.A. (2008) Prediction market accuracy in the long run. International Journal of Forecasting 24(2):285–300

Berg J.E., Rietz T.A. (2003) Prediction markets as decision support systems. Information Systems Frontiers 5(1):79–93

Boehmer-Christiansen S. A. (2016) Fuel for thought 27(1): Early November to end of 2015 (All dates refer to 2015/2016 unless stated otherwise). Energy & Environment, 27(1), 122–208

Bossaerts P., Frydman C., Ledyard J. (2013) The speed of information revelation and eventual price quality in markets with insiders: comparing two theories. Rev. Financer

Brier G.W. (1950) Verification of forecasts expressed in terms of probability. Monthly Weather Review 78:1–3

Budish, Eric B. and Cachon, Gerard P. and Kessler, Judd B. and Othman, Abraham (2015) Course Match: A Large-Scale Implementation of Approximate Competitive Equilibrium from Equal Incomes for Combinatorial Allocation. Chicago Booth Research Paper No. 15-28

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

Campbell D. (2016). What does the Paris Agreement actually do? *Energy & Environment*, *27*(8), 883–895

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

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

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

Chen Y., Kash I.A., Ruberry M., Shnayder V. (2014) Eliciting predictions and recommendations for decision making. ACM Transactions on Economics and Computation 2(2):6:1–6:27

Chen Y., Ruberry M., and Wortman V. J. (2012) Designing informative securities. In Proceedings of the 28th Conference on Uncertainty in Artificial Intelligence (UAI). 185–195

Christiansen J.D. (2007) Prediction Markets: Practical Experiments in Small Markets and Behaviours Observed, Journal of Prediction Markets, University of Buckingham Press, vol. 1(1), pages 17-41

Clemen R.T. (2002) Incentive contracts and strictly proper scoring rules TEST: An Official Journal of the Spanish Soceity of Statistics and Operations Research. Springer vol 11:195–217

Conradt, Larissa & Roper, Timothy (2005) Consensus decision making in animals. Trends in ecology & evolution. 20. 449-56. 10.1016/j.tree.2005.05.008

Couzin I.D., Krause J., Franks N.R., Levin S.A (2005) Effective leadership and decision-making in animal groups on the move. Nature.433:513–516

Cowgill B. and Zitzewitz E. (2015) Corporate Prediction Markets: Evidence from Google, Ford, and Firm X .The Review of Economic Studies 82 (4):1309–1341

Das, Rajarshi and Hanson, James & Kephart, Jeffrey & Tesauro, Gerald. (2001) Agent-Human Interactions in the Continuous Double Auction. IJCAI International Joint Conference on Artificial Intelligence

DeWispelare A.R., Herren L.T., Clemen R.T. (1995) The use of probability elicitation in the high-level nuclear waste regulation program. International Journal of Forecasting 11(1):5–24

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

Fama E.F. (1965) The Behavior of Stock-Market Prices. The Journal of Business, 38(1), 34–105

Fama E.F. (1970) Efficient Capital Markets: A Review of Theory and Empirical Work. Journal of Finance. 25 (2): 383–417

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

Gigerenzer G., Hertwig R., Van Den Broek E., Fasolo B. and Katsikopoulos K. 2005. A 30% chance of rain tomorrow: How does the public understand probabilistic weather forecasts? Risk Analysis, *25*, 623–629

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

Good I.J. (1952) Rational decisions. Journal of the Royal Statistical Society. Series B (Methodological) 14(1):107–114.

Graefe A. and Weinhardt C. (2008) Long-Term Forecasting with Prediction Markets - a Field Experiment on Applicability and Expert Confidence, Journal of Prediction Markets, 2, 71-92

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

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

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

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

Hayek F. (1945) The Use of Knowledge in Society. American Economic Review, 35

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

Horn M., Walsh K., Zhao M., Camargo S.J., Scoccimarro E., Murakami H., Wang H., Ballinger A., Kumar A., Shaevitz D.A., Jonas J.A., and Oouchi K. (2014) Tracking scheme dependence of simulated tropical cyclone response to idealized climate simulations. J. Climate, 27(24): 9197-9213

Iyer K., Johari R., and Moallemi C. (2010) Information aggregation in smooth markets. In Proceedings of the 11th ACM Conference on Electronic Commerce (EC). 199–205

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

Lee M.D. and Shi J. (2010) The Accuracy of Small-Group Estimation and the Wisdom of Crowds.

Manski C. (2006) Interpreting the predictions of prediction markets. Economic Letters 91(3):425-429

Maskin E. (2011) Commentary: Nash equilibrium and mechanism design. Games and Economic Behavior, Elsevier, vol. 71(1), pages 9-11

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

Morgheim, J. BP Climate Change (2000) Statement to the Senate committee on commerce, science, and transportation, full committee hearing on solutions to reduce greenhouse gas emissions

Muth J. (1961) Rational Expectations and the Theory of Price Movements. Econometrica

Navajas J., Niella T., Garbulsky G., Bahrami B. and Sigman M. (2018) Aggregated knowledge from a small number of debates outperforms the wisdom of large crowds. Nature Human Behaviour, 2(2), 126-132

Nyarko, Yaw and Schotter, Andrew (2001) An Experimental Study of Belief Learning Using Elicited Beliefs. Econometrica vol. 70, no. 3, 2002, pp. 971–1005. *JSTOR*

O'Carroll F.M. (1977) Subjective probabilities and short-term economic forecasts: An empirical investigation. Applied Statistics 26(3):269–278

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

Ortner G. (1988) Forecasting markets-an industrial application. Mimeo, Technical University of Vienna

Ostrovsky M. (2012) Information aggregation in dynamic markets with strategic traders. Econometrica 80(6): 2595–2648

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

Page L. (2011) The ability of markets to predict conditional probabilities : Evidence from the US Presidential Campaign. In Williams, L V (Ed.) Prediction markets: Theory and applications [Routledge International Studies in Money and Banking series]. Routledge, United Kingdom, pp. 123-136

Pennock D.M. (2004) A Dynamic Pari-Mutuel Market for Hedging, Wagering, and Information Aggregation, ACM Conference on Electronic Commerce, New York, NY, USA, May 17-20

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

Plott C.R. and Sunder S. (1988) Rational expectations and the aggregation of diverse information in laboratory security markets. Econometrica 56(5):1085-1118

Plott C.R., Wit J. and Yang W. (2003) Parimutuel betting markets as information aggregation devices: experimental results. Economic Theory 22: 311-351

Polk K., Adler C., Muller D. and Rechtman K. (2003) Early Intervention: Diversion and Youth Conferencing: A National Profile and Review of Current Approached to Diverting Juveniles from the Justice System. Canberra: National Crime Prevention, Australia

Roth, A. E. (2015) Who gets what and why: the new economics of matchmaking and market design . Boston: Mariner Books

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

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

Tetlock P.E. and Mellers B. (2011) Intelligent management of intelligence agencies: Beyond accountability ping-pong. American Psychologist 66(5) 542-554

Tetlock P.E. and Mellers B. (2014) Judging political judgment. Proceedings of National Academy of Sciences, 111 (32) 11574-11575

Tversky A. and Kahneman D. (1979) Prospect Theory: An Analysis of Decision Under Risk. Econometrica 47 263-266

Tziralis G. and Tatsiopoulos I. (2007) Prediction markets: An extended literature review. Journal of Prediction Markets 1(1):75–91

Vickrey W. (1961) Counter speculation, Auctions and Competitive Sealed Tenders. Journal of Finance, 16, 8—39

Wagman L. and 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

Winkler R.L. (1969) Scoring rules and the evaluation of probability assessors. Journal of the American Statistical Association 64(327):1073–1078

Wolfers J. and Zitzewitz E. (2004) Prediction markets. Journal of Economic Perspectives, 18(2), 107-126

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

Wunderlich F. and Memmert D. (2016) Analysis of the predictive qualities of betting odds and FIFA World Ranking: evidence from the 2006, 2010 and 2014 Football World Cups, Journal of Sports Sciences, 34:24, 2176-2184

Chapter 2: An Experimental Approach to Study Information Aggregation in Decision Markets

"On an important decision one rarely has 100% of the information needed for a good decision no matter how much one spends or how long one waits. And, if one waits too long, he has a different problem and has to start all over. This is the terrible dilemma of the hesitant decision-maker."

— **Robert K. Greenleaf, The Servant as Leader**

2.1 Introduction

Everyday decision-making involves the incorporation of complex information. Theory and experience indicate that the aggregation of such information in groups leads to better judgments than the information of any one member of the group (Surowiecki 2004). Thus, reaching an optimal decision often necessitates the integration of knowledge from multiple dispersed sources.

Many economic mechanisms have been designed to aid and guide decision making in groups. This includes prediction markets which aggregate information via market mechanisms for the primary purpose of predicting future outcomes. By allowing individuals to trade on assets with payoffs tied to future outcomes, the private beliefs of individuals can then be elicited and aggregated to provide informative forecasts about probabilities of future outcomes. For example, in 2011, to predict the election outcome of New Zealand, iPredict launched a set of three contracts that pay out $1 per owned asset if the elected Prime minister will be from Labour, National, or neither Labour or National. The price of such an asset can be interpreted as the probability assigned by the market participants to the corresponding outcomes.

Most real-world prediction markets employ an incentive structure that is equivalent to a zero-sum game for the forecasters, such as markets with double auctions to match different traders, or pari-mutuel markets. Such prediction markets rely on intrinsic motivations for participants to make predictions and reveal information. In the absence of such intrinsic motivations, trading cannot be expected. However, prediction markets can be designed to offer participants proper incentives to make accurate predictions (Hanson 2003). As described in Chapter 1, Hanson (2003) describes a market scoring rule which can be implemented as automated market maker (AMM) to incentivize traders to reveal the information they possess instantaneously by pricing it into the market.

By eliciting predictions about future events, prediction markets can be used as tools to guide decisions. A company may, for instance, guide investments depending on likely election outcomes to align with expected future policy changes. There are, however, instances where the probabilities of realized outcomes depend on an action taken by a decision-maker. A decision-maker in such a situation requires conditional forecasts, i.e., what happens if a particular action is taken. Such conditional forecasts are characteristic for decision markets.

Decision markets elicit forecasts that are conditional upon the executed decision (Chen et al. 2011, Teshner et al. 2017), and use these forecasts to inform decisions. In other words, they simultaneously help to "predict and decide'' the future. Recall the hiring example in Chapter 1, where a decision-maker creates two conditional prediction markets to know which developer would be more likely to complete an assigned project on time. If two prediction markets were run, then both developers would have to be hired to resolve the markets. However, in a decision market, *only one* developer is hired, and the performance of the other developer cannot be assessed. This element of conditional forecasting means that we do not witness all outcomes, because some decisions are not executed (Othman and Sandholm 2010). Since proper scoring rules rely on outcomes being observed, maintaining incentive compatibility for such a mechanism is difficult. To resolve this problem, Chen et al. (2014) introduced stochastic decision rules as a way to map incentivized forecasts and decisions. Stochastic decision rules are more complex than simple deterministic rules for participants to evaluate but implementing such a rule allows to design decision markets where the best strategy is to invest exactly according to one's beliefs.

In this chapter we describe the design of an experiment that aims to study how decision rules in decision markets affect information aggregation. We investigate whether the predictive performance of prediction markets can be matched by a decision market, and if so, whether it is sensitive to the decision rules used. To do so, we implement decision markets in a controlled laboratory setting with pre-established 'experts' that trade on markets employing different decision rules. An analogue of Plott and Sunders' (1988) prediction market experiment is used to provide an experimental demonstration for information aggregation in decision markets. Our market settings resemble potential real-world settings, where there are significant knowledge barriers, and inexperienced traders are the norm.

The rest of the chapter is structured as follows: in section 2.2 we explore related experiments, 2.3 outlines the design of the experiment and describes our recruitment procedures and payment, 2.4 quantifies our market errors and 2.5 concludes by summarizing the chapter.

2.2 Related Experimental Work

2.2.1 Prediction market studies

Several studies on prediction markets have used experiments to demonstrate that even with a limited pool of traders, aggregating information can be efficient. Christiansen (2007) created a set of small-scale prediction markets in order to compare market calibration, trading behaviour and participation incentives to those of large-scale markets, such as the Hollywood Stock Exchange and derivatives market on economic indicators (Pennock et al. 2001; Gürkaynak and Wolfers 2005).[13] Small prediction markets on the outcome of rowing regattas in the United Kingdom were designed for participants to predict the winners. Such a sport has significant knowledge barriers given that no news media regarding British rowing is broadcast outside of its national club and there is no forum for aggregating opinions. Such barriers reflect potential real-world problems for prediction markets. There were no experienced traders amongst the subject pool of participants, and this resembled real-world prediction markets, where participants that have relevant information about a future outcome may not have experience in betting or derivatives trading. Christiansen's study (2007) found that information aggregation in such small-scale markets were efficient and having as few as sixteen traders can generate reliable information.

Plott and Sunder (1988) pioneered what has become the classic information aggregation experiment. In their experiments, the true state of the world can be one of three possible mutually exclusive outcomes; X, Y or Z. For each state other than the

[13] The Economic Derivatives market are large-scale markets that were run by Goldman Sachs and Deutsche Bank. They are directly linked to macroeconomic outcomes and are occasionally run as pari-mutuel markets. Gürkaynak and Wolfers (2005) provide a more detailed description of their structure.

true state, half of the traders are informed of the false state. For example, if the true state is Y, half of the traders are informed that it could either be X or Y (i.e., not Z), and the other half are informed that it could be either Y or Z (i.e., not X). Thus, traders have heterogeneous information that, if aggregated efficiently, is enough to fully reveal the true state. The results from their trading rounds in the experimental study showed that market prices frequently converge to represent the true state of the world.

Page and Siemroth (2017) explore information acquisition in prediction markets in a laboratory setting by allowing participants to trade in markets on types of urns. Participants had the option of acquiring several pieces of information, provided as draws from an urn, at a cost and trade according to their beliefs about the true state of nature. The authors report that bidders are more likely to obtain costly information if they have bigger cash and asset endowments, if their current knowledge is ambiguous, and if they are less risk averse. They also suggest that the traders' inclination to over-acquire information may be part of the reason why prediction markets' prices are informative. Their findings show how prediction markets can provide incentives for participants to obtain information, and demonstrates that using urns is a well-suited approach for representing events.

2.2.2 Decision market studies

The first experimental insights on decision markets were published in a PhD thesis by Forsell (2016). The author aimed to evaluate whether the prediction accuracy of decision markets provided more accurate predictions than those of a prediction market using different combinations of scoring rules. Participants traded in laboratory markets with three different incentive structures which allowed comparing the accuracy of aggregated private information. Trading was designed to forecast a continuous variable, using a set of 5 contracts on non-overlapping intervals in which the continuous variable could fall. He compared two types of decision markets, with proper (i.e., stochastic) and improper (i.e., deterministic) decision rules, and additionally included strictly proper prediction markets as control. A proper decision market constituted of using strictly proper scoring rules whilst an improper market was one that did not. Their results show that irrespective of the decision rule, decision markets performed worse than a prediction market. There was no statistically significant difference in the accuracies of the proper and improper decision markets.

Teschner et al. (2017) conducted decision market experiments on the Amazon Mechanical Turk platform. They provided the first experimental investigation of the influence of manipulation on decision quality, with manipulation being defined as deviations from ‘’truth-telling’’, which can be incentivized under improper market design. Two online experiments were designed to examine the degree and impact of manipulation on forecasting accuracy in decision markets. Participants were presented with a series of decision tasks, represented by urns, and asked to predict the outcomes. In the first experiment, subjects interacted with an automated market maker. In the second experiment they interacted with the market maker and one additional trader. In each experiment, four treatments were implemented that examined theoretical predictions about decision quality. These four treatments corresponded to different decision rules, allowing the authors to study under which decision rules subjects manipulate the market. The results of their experiments demonstrated that manipulation could occur under different incentive-schematic settings. Manipulation occurred in the first experiment but not in the second as the entry of another trader added risk to the manipulation strategy.

Until recently, and to the best of our knowledge, these two experiments represent the only demonstrations of different aspects in decision markets. We harness similar approaches but with different market designs, objectives, and analyses. We use a simpler design compared to Forsell (2016) and Teschner et al. (2017). While Teschner et al.'s (2017) implementation on Amazon Mechanical Turk did not allow for participants to trade repeatedly and simultaneously with other market participants as is typical in many real-world market settings, our experiment allows individuals to trade regularly and concurrently with other market players. Meanwhile, similarly to Forsell (2016), we compare the performance of decision markets to prediction markets to allow for a better understanding of the functioning of decision markets. We particularly focus on a key feature of markets, namely information aggregation. As such, we follow an analogue of Plott and Sunders's (1988) prediction market experiment, with decision market approaches similar to those used by Forsell (2016) and Teschner et al. (2017) to provide experimental insights about the functioning of decision markets.

2.2.3 Experimental considerations

Following the findings and recommendations from Forsell (2016) and Teschner et al. (2017), we note the importance of implementing a simple design framework to cultivate a successful market environment.

As will be described in detail in Section 2.3, we provide participants with opportunities to gain experience over the course of our experiment by introducing a prediction market as a baseline, several practice rounds and a summary screen after each practice round. Thereby participants will be able to familiarize themselves with the trading interface and have more confidence in the trades being made whilst making them "experts" in the market.

To mitigate the complexity of the experiment, we make sure that the experimental stages are designed to be as similar as possible, more vernacular wording is used in the instructions, separate instructions are verbally explained before each stage and presented via a presentation whilst any questions are answered publicly.[14] Furthermore, to minimize any misunderstandings, prior to trading, participants are tested with key questions about the markets and the designated treatment (given in *Appendix A*).[15]

Forsell (2016) highlights the importance of traders preferring clearer, binary market structures. We take into consideration the implications of such alternative mechanism designs and as a result, we use asset types that are less complex. Our settings are closely matched to those of binary lotteries to allow participants to easily assess the incentive structure. By taking the author's recommendations as guidance fosters a more rapid learning environment and reinforces a more robust market design.

2.3 Experimental Design and Procedures

We aim to contribute to the literature exploring information aggregation in decision markets given the general paucity of work in this area as suggested by the discussion above in Section 2.2. Forsell (2016) explored the predictive accuracy of decision markets under different scoring rules in comparison to prediction markets, but due to the high complexity of such markets, their market performance was not as good as

[14] Instruction slides are given in Appendices B and C.
[15] Contrary to Teschner et al. (2017), we do not drop out participants who don't get a perfect score in the control questions. Instead, before beginning a stage, we ensure that the correct response to each control question is obtained.

anticipated. Teschner et al. (2017) focused more on the influence of manipulation in decision markets with proper and improper decision rules. However, their experiments were not designed specifically to investigate information aggregation and allowed each participant to change asset prices only once. Our objective is to provide a simple proof-of-principle for aggregation of private information within decision markets. Do properly incentivized decision markets aggregate information as efficiently as prediction markets? And does the properness of the decision rule affect market accuracy?

To address this gap, it is essential to implement a simple, controlled experimental environment. Following a similar market approach to the one used by Teschner et al. (2017), we adapt Plott and Sunders' (1988) experiment regarding information aggregation in prediction markets to a conditional forecasting scenario. Here, one of two possible mutually exclusive outcomes will be realized, and this realization depends on a decision between two mutually exclusive actions. Moreover, similar to Page and Siemroth (2017), we use the experiment to explore relations between market performance and behavioural attributes of individual traders but instead of adopting a Dohmen et al. (2011) approach, we use the popular Multiple Price List (MPL) by Holt and Laury (2002) as it provides simple incentives for truthful preference revelation and has repeatedly been proven effective in explaining behavior in market settings (Fellner and Maciejovsky 2007).

Our experiment resembles a situation where a decision-maker has a preference regarding the outcomes and needs to elicit which of the two actions is the best, i.e., most likely to result in the preferred outcome. As such, a market is set up to incentivize 'experts' to reveal and aggregate their information so that the decision-maker can make an informed decision. Participants hold private pieces of information regarding how actions translate into outcomes. They trade assets similar to Arrow-Debreu securities (binary options), which are frequently used in prediction markets and feature a payoff of 1 if a specific outcome is realized, and a payoff of 0 if a different outcome occurs. As Plott and Sunder's experiment demonstrated accurate information aggregation in prediction markets in the absence of conditional dependency, we are using a prediction market setting as a control.

We use a three-stage laboratory-based experiment to study how well decision markets work in aggregating private information.

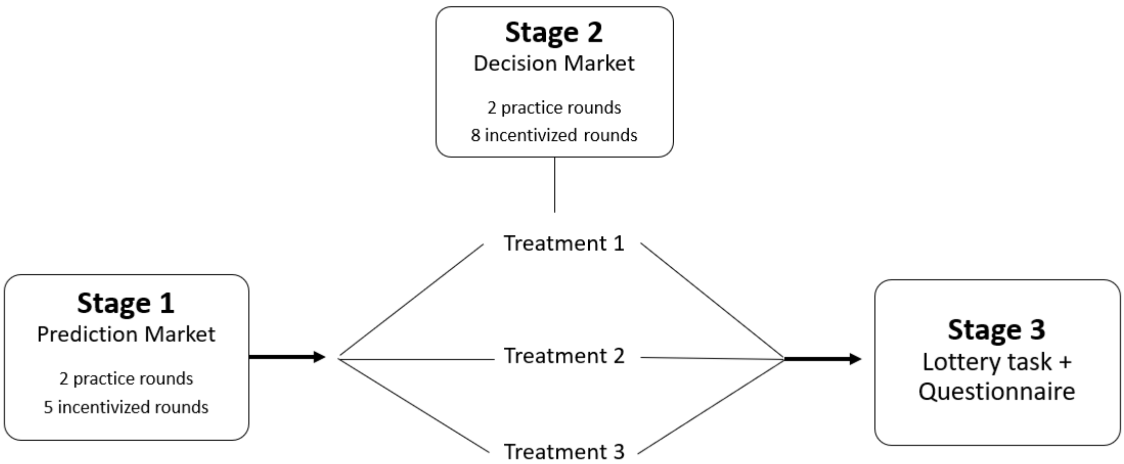


Figure 2.1 Experimental design workflow

Participants trade in these markets to predict uncertain outcomes of events, which are represented by urns. In Stage 1, we use a prediction market setting with two unconditional outcomes to capture how well information aggregation works in such a simple baseline treatment. This stage is the same for all participants (see Figure 2.1), has two practice rounds, followed by five incentivized rounds, and provides an opportunity for participants to get familiar with the trading interface. In Stage 2, we expand this approach to a decision market setting for two actions and two conditional outcomes for each action We use three different treatments (further explained in

Section 2.3.3) in this stage, two of which are decision markets with different decision rules, and one is a prediction market that matches the market design complexity of the decision markets. As such, we have three different market designs, and they differ in the decision rule and how the payoffs are being made, which is related to the decision rule. Stage 3 consists of a post-experimental questionnaire to further our understanding of participants' trading behavior. We implement a between subjects treatment where each participant experiences only one of the three treatments in Stage 2. The treatment a participant experiences in Stage 2 depends on which experimental session she chose to sign up to.

Details about the urns and the information structure is provided in section 2.3.1. Section 2.3.2 outlines details about Stage 1, 2.3.3 explains Stage 2 and the corresponding treatments, Section 2.3.4 details Stage 3 and 2.3.5 describes the recruitment and payment procedures.

2.3.1 Urns, private information, market design

Urns: Similar to experiments conducted by Anderson and Holt (1997), followed by Page and Siemroth (2017), Teschner et al. (2017) and Halim et al. (2019), we use urns as a model to generate private information about uncertain outcomes. Throughout our experiments there are two types of urns, a Red Type urn and a Blue Type urn. These types represent different outcomes. Each urn contains 10 balls that can be red or blue. The content of the urn depends on its Type. A Red Type urn contains 8 red balls and 2 blue balls. A Blue Type urn contains 2 red balls and 8 blue balls. These compositions are known to participants. Whether an urn in the experiments is of a Blue Type or a Red Type is selected randomly at a 50% chance each.

Private information: Before trading, each participant receives information about the type of urn in the form of one ball draw. Unlike in the study of Page and Siemroth (2017), participants only receive one signal. Each draw is the private information of one subject. When participants receive draws from the same urn, the draws are sampled randomly (like in the experiments by Oprea et al. 2007, Healy et al. 2010, Jian and Sami 2012, Deck et al. 2013), and independently with replacement. Moreover, they receive their signals simultaneously prior to each trading round. The private signal remains visible to the participants throughout the trading round and participants do not know anyone else's private signal, only their own.

Trading: Once the participants received their signal, they enter a market to trade assets. They are endowed with 20 e-dollars per round to buy and sell the assets. The value of these assets depends on the type of the urn. While the type is unknown to the participants, the private signal provides them with partial information about the type of the urn, which in turn is expected to affect their valuation of the assets. The private signal provides them with partial information about the type of the urn, which in turn is expected to affect their valuation of the assets. The e-dollars have no value except through buying assets in the respective markets. This fosters a positive trading environment and prevents participants from keeping their e-dollars to ensure a minimum payment. Any remaining balance after a trading round is set to zero . Trading is implemented with an automated market maker that uses a logarithmic scoring rule to update prices (Hanson 2003).[16] There is no communication or other interaction with other traders, except through the market (unlike Healy et al. 2010 or Jian and Sami 2012). This ensures that the information aggregation is solely achieved through the market rather than other communication channels. Even though participants do not directly trade with each other, their decisions affect asset prices and thus individual

[16] Refer to Chapter 1 Section 1.6.

payoffs depend on their own decisions and other participants' decisions. Each trading round lasts one minute and participants' pairings change every round throughout the whole experiment.

2.3.2 Stage 1: Prediction market with one urn

In each round in Stage 1, there is one urn. The type of the urn is unknown to the participants. It could be either of Red Type or of Blue Type. After receiving a draw from the urn, participants trade in a prediction market with two assets: one with a payoff of 1 e-dollar per Red Type asset if the urn is of a Red Type; and one with a payoff of 1 e-dollar per Blue Type asset if the urn is of a Blue Type. Thus, the task of participants it to estimate how likely the urn is a Red or Blue Type, and trade accordingly in the market.

For each round, the participants in a session are randomly split into groups with two participants each. Two is the minimal number of participants to observe information aggregation. Since our unit of observation to examine market efficiency is the individual market, participants are paired into groups of two to create a large number of small but independent markets. Similarly, Forsell (2016) used six participants per group; while Teschner et al. (2017) used single traders in one treatment and expanded to multiple traders in the second (as they focused on incentive-compatibility rather than information aggregation). Plott and Sunder, on the other hand, used more participants, and provided several participants the same piece of information. As a result, one would expect that information aggregation in their experiment to be more robust.

The two participants in a group receive signals from the same urn and trade in the same market. Thus there are two independently drawn signals (with replacement) for each urn, and two participants per market. Throughout the trading phase, subjects are able to see their private signal, number of e-dollars, number of invested assets, a graph with past and current prices of the assets, and the remaining time before market closure. A screenshot of the trading interface for Stage 1 is shown in Figure 2.2.

Figure 2.2 Screenshot of trading interface for Stage 1

At the end of the experiment, one of the five rounds in Stage 1 is randomly selected by the computer, the type of the urn for that round is revealed, and subjects are paid depending on the number of correct assets they hold in that selected round. For instance, if the urn is revealed to be of a Red Type, participants receive a payoff of $1 per Red Type assets and if the urn is of a Blue Type, they receive a payoff of $1 for each Blue Type assets. A summary screen is also displayed after each practice round and decisions from these rounds are not counted towards their final earnings.

2.3.3 Stage 2: Decision market with two urns

In each round in Stage 2, there are two urns: Urn 1 and Urn 2. Each urn could be of Red Type or of Blue Type. Participants receive a draw from one of the two urns. After receiving a draw from the urn, participants trade in a market with four assets. Two of the assets (Urn 1- Red Type and Urn 1-Blue Type) have payoffs that depend on the type of Urn 1, and two of the assets (Urn 2-Red Type and Urn 2-Blue Type) have payoffs that depend on the type of Urn 2. Despite their private signals being drawn from only one urn, they still can trade on both urns in the market. The payoffs of the assets depend on the settings and is explained further below.

In this stage, we expand our group size such that participants are split up into groups of four each. In each group, two participants receive an independently drawn signal each from Urn 1, and the other two participants an independently drawn signal from Urn 2. Thus, as in Stage 1 there are two independently drawn signals for each urn. Thus, although there are now two urns and four participants per market, there are still two independently drawn signals for each urn just as in Stage 1.

Figure 2.3 shows the trading interface for Stage 2. Subjects can see their private signal, the number of e-dollars and invested assets, graphs showing past and current prices of all four assets and the remaining time before market closure.

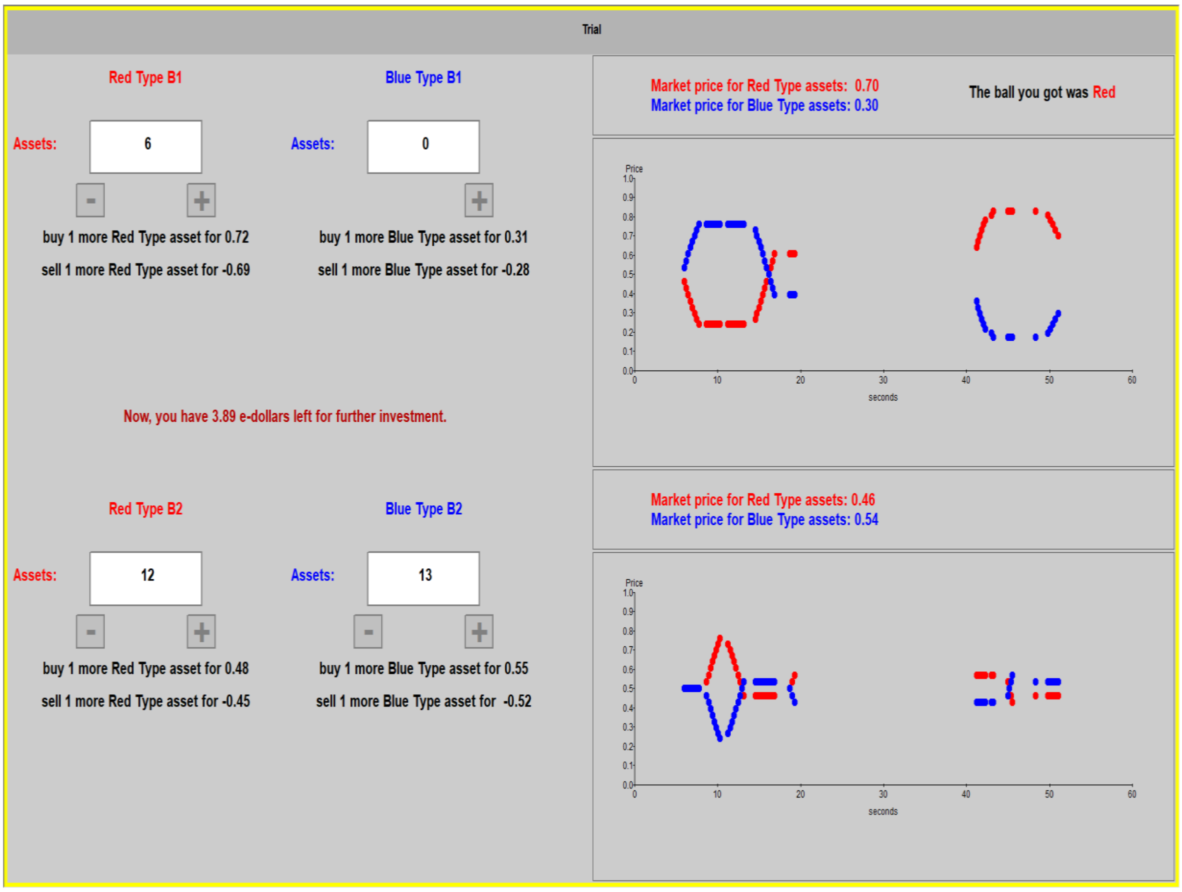


Figure 2.3 Screenshot of trading interface for Stage 2. Top-half of the screen represents the market for Urn 1 and the bottom-half market for Urn 2

Apart from the number of urns, the main difference between Stage 1 and Stage 2 is that Stage 1 employs no decision rule (since no decision needs to be made) while in Stage 2 there are three different treatments, each with a different decision rule. The decision rules differ in how asset prices determine which of the urns is being revealed, and in the payoffs paid for the assets.

I. Decision Rule 1: **Two Prediction Markets** ***(PM2)***

The Type of both urns are revealed, and participants are paid depending on the true outcome. Participants receive $1 for each correct asset. All other assets will be worth $0. For instance, if Urn 1 is of Red Type, and Urn 2 is of Blue Type, participants will receive $1 for each ‘Red Type B1’ asset and for each ‘Blue Type B2’ asset, and the other two assets will have a payoff of $0. This treatment is equivalent to two simultaneous prediction markets.

II. Decision Rule 2: **Deterministic *(DMD)***

This treatment represents a decision market where one of the urns is selected to be revealed based on the final market prices: the urn revealed is the one that the final market price indicates is most likely to be a Red Type. The type of the other urn **is not** revealed. In other words, if the price for 'Red Type B1' is larger than the price for 'Red Type B2', Urn 1 is selected, and its type can be observed. Otherwise, Urn 2 will be selected. A payoff will only be received for assets on the type of the revealed urn; all assets for the other urn will be worth $0. For instance, if Urn 1 is selected and is of Red Type, participants will receive $1 for each 'Red Type B1' asset, and the other three assets will have a payoff of $0. This treatment represents a decision market with a deterministic decision rule and is not incentive compatible.

III. Decision Rule 3: **Stochastic *(DMS)***

This treatment represents a decision market where one urn is stochastically selected to be revealed. The probabilities depend on the final prices in the market. The urn which according to the market price is most likely of a Red type is selected at a probability of 75%. The other urn is selected at a probability of 25%. The type of the selected urn is revealed, and a payoff is only received for owned assets on the type of the revealed urn. All assets for the other urn will be worth $0. The payoff for the paying assets depends on the probability at which the urn is selected; if an urn is selected at p, the payoff will be $1/p. All other assets in the other market are voided.

If, for instance the price for ‘Red Type B1’ is larger than the price for ‘Red Type B2’, Urn 1 is selected at p=0.75, and Urn 2 is selected at p=0.25. If Urn 1 is selected and is of Red Type, participants will receive \$1/0.75=\$1.33 for each ‘Red Type B1’ asset, and the other three assets will have a payoff of \$0. If Urn 2 is selected, and is of a Red Type, participants will receive \$4 for each ‘Red Type B2’ asset, and the other three assets will have a payoff of \$0. This treatment represents a decision market with a stochastic decision rule and is incentive compatible. Table 2.1 provides a quick summary behind the reasons of using such decision rules.[17]

Table 2.1 Summary of the decision rules used in the experiment

Treatment	Incentive-compatible	Reason
Two prediction markets (PM2)	Yes	A control to determine how much efficiency we lose from having two markets at the same time (from PM1). If performance is not as efficient as our baseline, this would be the cost of employing two simultaneous prediction markets.
Deterministic (DMD)	No	Not incentive compatible. Teschner et al. (2017) showed that deterministic decision rules work better with less noise than probabilistic ones.
Stochastic (DMS)	Yes	This setting follows Chen et al.’s (2014) description and is incentive compatible. Comparison with the other settings will establish the importance of proper decision rules.

At the end of the experiment, one of the eight rounds in Stage 2 is randomly selected by the computer and depending on the decision rule the type of either Urn 1 or Urn 2,

[17]A more elaborate discussion is provided is Chapter 1.

or both (in treatment PM2) is revealed, and subjects are paid depending on the number of correct assets they hold. As for Stage 1, a summary screen is also displayed after each practice round and decisions from these rounds are not counted towards their final earnings.

2.3.4 Stage 3: Demographic questionnaire and risk lottery

After all decision-making tasks have been completed in Stage 1 and Stage 2, participants are presented with a questionnaire about demographic information (given in *Appendix D & E*) and a task to measure individual risk attitudes. Risk attitudes are elicited by adopting the Holt and Laury (2002) Multiple Price List (MPL) where an ordered array of 10 binary lottery choices is presented to participants (given in *Appendix E*). Choices from the MPL tasks contribute towards their final earnings (maximum of 3.85 New Zealand Dollars). A summary screen displaying participants' decisions, the true outcomes of the selected rounds and its corresponding payoffs are displayed at the end of the experiment.

2.3.5 Recruitment and payment

A total of 96 subjects participated in our twelve experimental sessions. Each participant took part once in one of our experimental treatments: 1) PM1 and PM2, 2) PM1 and DMD and 3) PM1 and DMS. Four sessions per treatment were conducted. Sessions were conducted at the University of Auckland DECIDE laboratory, and participants were recruited via the ORSEE system (Greiner 2004). Our market interface and the survey were implemented in z-Tree (Fischbacher 2007).[18] Sessions lasted a

[18] Similar market experiments used Mechanical Turk (MTurk) but due to the necessity of maintaining group compositions and the complexity of understanding such markets, using z-Tree in-the-lab was more robust.

maximum of 90 minutes including approximately ten to fifteen minutes time for the instructions and ten minutes for pay-outs.

As described above, participant's payoffs are linked to their individual performance, final holdings and treatment used. One round from each market Stage (Stage 1 and Stage 2) is selected randomly and independently for payment. This randomly chosen decision is called the random problem selection (RPS) mechanism (Allais 1953). With this mechanism subjects cannot receive payoffs for all decisions, and therefore have no incentive to favor one choice in all decision problems. As such, participants will be inclined to perform truthfully in line with their information. Apart from the \$10 show-up fee, participants earned a minimum of \$25 if they decided to stay until the end of the experiment and a maximum of \$50 depending on their performance throughout the stages of the experiment. Choices in the practice rounds did not contribute towards their final earnings.

2.4 Quantifying Market Errors

We analyse information aggregation in our market settings by examining forecasting errors per treatment. The market price is assumed to reflect the aggregate prediction of the probability of the respective urn to be a Red or Blue Type. A higher market price for Blue Type assets tend to indicate that participants believe that the true state of an urn is of a Blue Type, whilst a higher market price for Red Type assets, tend to indicate a Red Type urn.

Forecasting errors capture the difference between the final market price, P^{M}_{i} of a market i, and the correctly aggregated Bayesian probability P^{B}_{i} given the signals shown to participants. For each urn, two signals are independently drawn and revealed to different participants. The Bayesian probability P^B_i for a urn to be of Red Type is $P(r|R,R) =$ 0.941 if both balls are red, $P(r|B,B) =$ 0.059 if both balls are blue and $P(r|R,B) =$ 0.5 if one ball is red and one is blue. For a more detailed demonstration of correct aggregation, see *Appendix F*.

Put into context, a participant choosing to invest 15 e-dollars of her endowment (out of 20 e-dollars) on Red Type assets would indicate her strong belief that an urn would be of a Red Type and thus increase the asset price from its starting value of 0.5. If the Red Type asset costs $0.65, it indicates that the market participants think there is a 65% chance that the urn will be of a Red Type. As such, an asset "traded" at $0.65 can be interpreted as a forecast for that outcome to occur at a 65% chance.

A market price, e.g., of $0.65 for a Red Type asset, indicates to a market participant what other market participants believe. A trader who receives a signal will move the price into a specific direction. The other trader may either learn from this price change and their own signal to aggregate the available information appropriately. Using this type of communication, the method in which information is revealed in the market allows all market participants to conclude on the signal and then apply the right Bayesian aggregation to guide their own evaluation. As a result, asset prices can be expected to follow Bayesian aggregation, which is why we use it as a benchmark for price accuracy.

To quantify information aggregation in the markets, we use the following three quantities:

1. **Error:** the difference between the final prices in the market and the correct Bayesian probability, $err_i = P^M_i - P^B_i$

2. **Absolute Error:** the absolute difference between the final market prices and the correct Bayesian probability $errabs_i = abs(P^M_i - P^B_i)$

3. **Absolute Logit Error:** the difference of the log odds of the final market prices and the correct Bayesian probability $logiterrabs_i = abs(logit\ P^M_i - logit\ P^B_i)$. This transformation increases errors for probabilities that are closer to zero or one and makes the distribution of errors more similar to a normal distribution. We use this alternative method as a robustness test to measure error where the errors are being amplified for extreme probabilities.

2.5 Summary

Decision markets gather experts to predict the effects of each of a set of possible actions (Chen et al. 2011). It uses the market forecasts to assign a probability to each action, and then uses a decision scoring rule to reward experts for accurate predictions. In the experiments described in this chapter, we simulate a situation where a decision-maker must make a choice over possible decisions to execute, where the possible decisions will impact the probability of achieving a desired outcome. We examine three possible decision rules for experts to help us identify if information is aggregated within these markets. We use a prediction market as a baseline to assess the performance of the more complicated decision markets mechanisms.

References

Allais M. (1953) Le comportement de l'homme rationnel devant le 67isqué, critique des postulats et axioms de l'ecole Americaine. Econometrica 21: 503–46

Anderson L. R. and Holt, C.A. (1997) Information Cascades in the Laboratory. The American Economic Review, 87(5), 847–862.

Berg J.E. and Rietz T.A. (2003) Prediction markets as decision support systems. Information Systems Frontiers 5(1):79–93

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

Chen Y., Kash I.A., Ruberry M. and Shnayder V. (2014) Eliciting predictions and recommendations for decision making. ACM Transactions on Economics and Computation 2(2):6:1–6:27

Christiansen J.D. (2007) Prediction Markets: Practical Experiments in Small Markets and Behaviours Observed. Journal of Prediction Markets, University of Buckingham Press, 1(1): 17-41

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

Dohmen T., Falk A., Huffman D., Sunde U., Schupp, J. and Wagner G.G. (2011) Individual risk attitudes: measurement, determinants, and behavioral consequences. Journal European Economics Association 9, 522–550

Fellner G. and Maciejovsky B. (2007) Risk attitude and market behavior: Evidence from experimental asset markets. Journal of Economic Psychology 28(3), 338-350

Fischbacher U. (2007) z-Tree: Zurich toolbox for ready-made economic experiments. Experimental Economics 10, 171–178

Forsell E. (2016) Experimental Testing of Old and New Hypotheses in Economics, Dissertation for the Degree of Doctor of Philosophy

Greiner B. (2004) An online recruitment system for economic experiments. http://mpra.ub.uni-muenchen.de/13513/1/MPRA_paper_13513.pdf

Gürkaynak R.S. and Wolfers J. (2005) Macroeconomic derivatives: an initial analysis of market-based macro forecasts, uncertainty and risk, IZA Discussion Papers, No. 1899, Institute for the Study of Labor (IZA), Bonn 11-64

Halim E., Riyanto Y.E. and Roy N. (2019) Costly information acquisition, social networks, and asset prices: experimental evidence. The Journal of Finance, 74(4), 1975-2010

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

Healy P.J., Linardi S., Lowery J.R. and Ledyard J.O. (2010) Prediction Markets: Alternative Mechanisms for Complex Environments with Few Traders. Management Science 56(11):1977–1996

Holt C.A. and Laury, S.K. (2002) Risk Aversion and Incentive Effects. American Economic Review, 92 (5): 1644-1655

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

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

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

Page L. and Siemroth C. (2017) An Experimental Analysis of Information Acquisition in Prediction Markets. Games and Economic Behavior, 101, 354-378

Pennock D., Lawrence S., Nielsen F. and Giles C. (2001) Extracting Collective Probabilistic Forecasts from Web Games. Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining

Plott C.R. and Sunder S. (1988) Rational expectations and the aggregation of diverse information in laboratory security markets. Econometrica 56(5):1085-1118

Roth A. E. (2015) Is Experimental Economics Living Up to Its Promise? In Fréchette, Guillaume R. and Andrew Schotter (editors) Handbook of Experimental Economic Methodology, Chapter 1, pp. 13-40, Oxford University Press

Surowiecki J. (2004) The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business, economies, societies, and nations. Doubleday & Co.

Teschner F., Rothschild D. and Gimpel H. (2017) Manipulation in Conditional Decision Markets. Group Decision Negotiation 26, 953–971

Chapter 3: Market Performance and its Dependence on Decision Rules

''The lone wolf dies but the pack survives.''-
Sansa Stark from The Game of Thrones

3.1 Introduction

In this chapter we describe the experimental results from our decision market experiment as outlined in Chapter 2. In particular, we analyze how efficiently decision markets aggregate information, and the extent to which their predictive performance matches that of a prediction market.

Since we know the signals provided to the participants and thus the correctly aggregated probabilities for the outcomes, information aggregation can be assessed for each market once completed by comparing final market prices with correctly aggregated probabilities. As described in Chapter 2 section 2.4, we quantify information aggregation in our market analysis by using absolute error, and as a robustness test, absolute error of logit transformed probabilities.

We investigate our research question by implementing a set of ordinary least-square (OLS) models with prices and errors as dependent variables, and a range of experimental parameters as independent variables. We address two main hypotheses:

Hypothesis 1: Do final market prices depend on private signals?
Hypothesis 2: Do market errors depend on the decision rule used?

In this section we provide a description of the variables used for our analysis in table 3.1 and the summary statistics of participants in table 3.2. Section 3.2 describes the results of Stage 1, which includes the test of the first hypothesis, i.e., the dependency of final market prices on private signals, in table 3.3. An analysis of the impact of experimental parameters on market accuracy is provided in tables 3.4 and 3.5. Section 3.3 describes the results of Stage 2 and includes a further test of the first hypothesis in table 3.6. The results for the second main hypothesis, i.e., whether the dependency of market error depends on the decision rule used, are given in tables 3.7 and 3.8. A follow up analysis from our findings for the second main hypothesis is given in tables 3.9, 3.10 and 3.11 where we investigate the differences in prices and errors in the different decision market settings for different signal constellations. We conclude and discuss the results of this chapter in section 3.4.

Table 3.1 Description of variables

Variable name	Description
Signals	The constellation of independently drawn signals provided to the two participants, where each participant only sees one ball. Red, Red (RR): the two independently drawn private signals from the urn are both red balls. Red, Blue (RB): the signals are one red and one blue ball. Blue, Blue (BB): the two signals are both blue balls.
Round	Round within a stage at which the market occurred. Stage 1 has five rounds with a value of 1-5 and Stage 2 eight rounds with a value of 1-8.
SessionID	A total of twelve sessions were conducted for all treatments throughout our experiment. This variable takes a value of 1-12, with 1 representing the first session and 12 the last session. Including the session number allows us to test for any session-effects that might arise. It helps us identify if any dependent variables (such as market accuracy) change as the experiment progresses (since the verbal instructions change as the experimenters gained experience).
Urn	This variable corresponds to an urn within a market. In a market, participants trade on the type of the urn(s). Stage 1 always has a single urn (labelled 1) and is thus not included in our statistical models. Stage 2 has two urns; one for each

	of the two markets where subjects participated. Market 1 trades on the type of urn 1 and market 2 trades on the type of urn 2. The top half of our trading interface always shows the market for Urn 1 and the bottom half the market for Urn 2 (see Chapter 2 Figure 2.4). As such, this variable takes the value of 1 or 2. We include this variable in our statistical models as the difference in the experimental design could affect how price is aggregated i.e., some participants could pay more attention to the urn at the top half of the trading interface rather than the bottom half. Including this variable would help us identify whether this design feature affects the outcomes we observe.
Price	The final market prices of Red Type assets in the market.
Setting	In Stage 2 there are three different treatments, each with a different decision rule **PM2**: Two Prediction Markets **DMD:** Deterministic **DMS:** Stochastic
Error	The difference between the final prices in the market and the correct Bayesian probability (described in Chapter 2 section 2.4)
Absolute Error	The absolute difference between the final market prices and the correct Bayesian probability (see section 2.4)
Absolute Logit Error	The difference of the log odds of the final market prices and the log odds of the correct Bayesian probability (see section 2.4)

Participants

Table 3.2 Descriptive characterstics of participants

Participants Characteristics	Count	Percentage
Gender		
Male	45	46.88
Female	51	53.12
Ethnicity		
New Zealand/Māori	3	3.12
Pakeha/ European	27	28.12
Pacific Islander	5	5.21
Asian	52	54.17
Middle Eastern	4	4.16
Indian	12	12.5
Other Ethnic/Group	2	2.08
School		
Business	55	57.29
Engineering	16	16.67
Sciences	23	23.96
Arts	16	16.67
Other	15	15.62
Taken prior statistics course		
Yes	78	81.25
No	18	18.75

Table 3.2 presents descriptive statistics of the selected participants. In total, 96 subjects participated in our 12 sessional treatments in August and October 2020. Due to random assignment amongst all our treatments, participants characteristics are expected to be balanced. The collected data consists of 240 markets for our prediction market setting (12 sessions × 4 Groups × 5 Rounds) and 384 markets (12 sessions × 2 Groups × 8 Rounds × 2 urns) for our decision markets setting. Our sample distribution was 56% females, with an approximate of 68% were between the ages of 18 and 22 (with mean age of 22 and SD 3.4). A total of 81% participants have taken a statistical course and 57% of our subject pool came from the Business School cohort. Our majority sample came from an Asian ethnicity with a total of 54%.

3.2 Results from Stage 1

In Stage 1, participants traded in a simple prediction market in groups of two. Each of the two participants received one independently drawn signal, and thus there are three possible signal constellations: RR, RB, and BB (see Table 3.1).

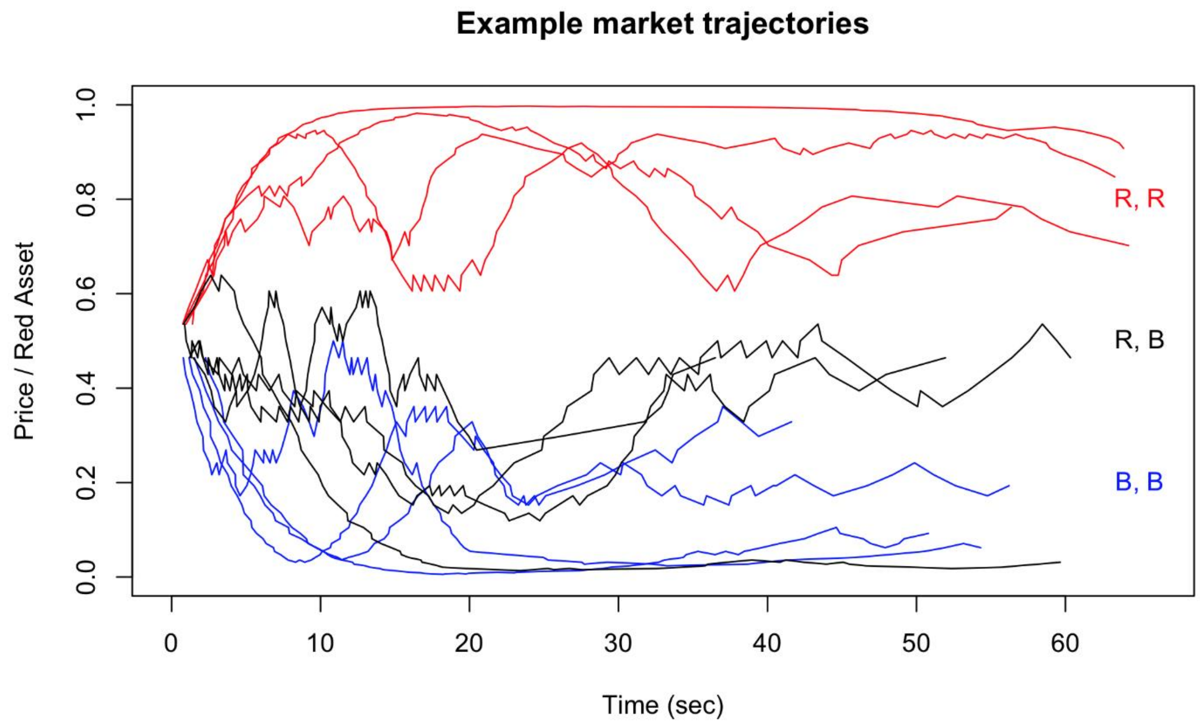


Figure 3.1 Example of price market trajectories for the first three rounds for Session 1

Figure 3.1 shows examples of the market trajectories for four groups of participants from the first session for the first three rounds and all four groups. Each round differs in the signals, and the three different signal constellations are RR, RB, or BB. The red trajectories correspond to constellations where both traders receive red signals, black trajectories correspond to constellations where one trader receives a red ball and the other trader a blue ball. Blue trajectories correspond to both traders receiving two blue signals. Recall from Chapter 2 that the Bayesian probability of a market i for an urn to be of Red Type given the possible combinations of private signals is $P(r|R,R) = 0.94$ if both balls are red, $P(r|B,B) = 0.06$ if both balls are blue and $P(r|R,B) = 0.5$ if one ball is red and one is blue. One can observe that most market price trajectories move towards the correct probabilities. An exception is one group in the RB category that moves far away from the correct probability towards a very low price for red assets. Figure 3.1 illustrates that markets largely perform the way we expect them to, but there is variation on how well markets aggregate the information.

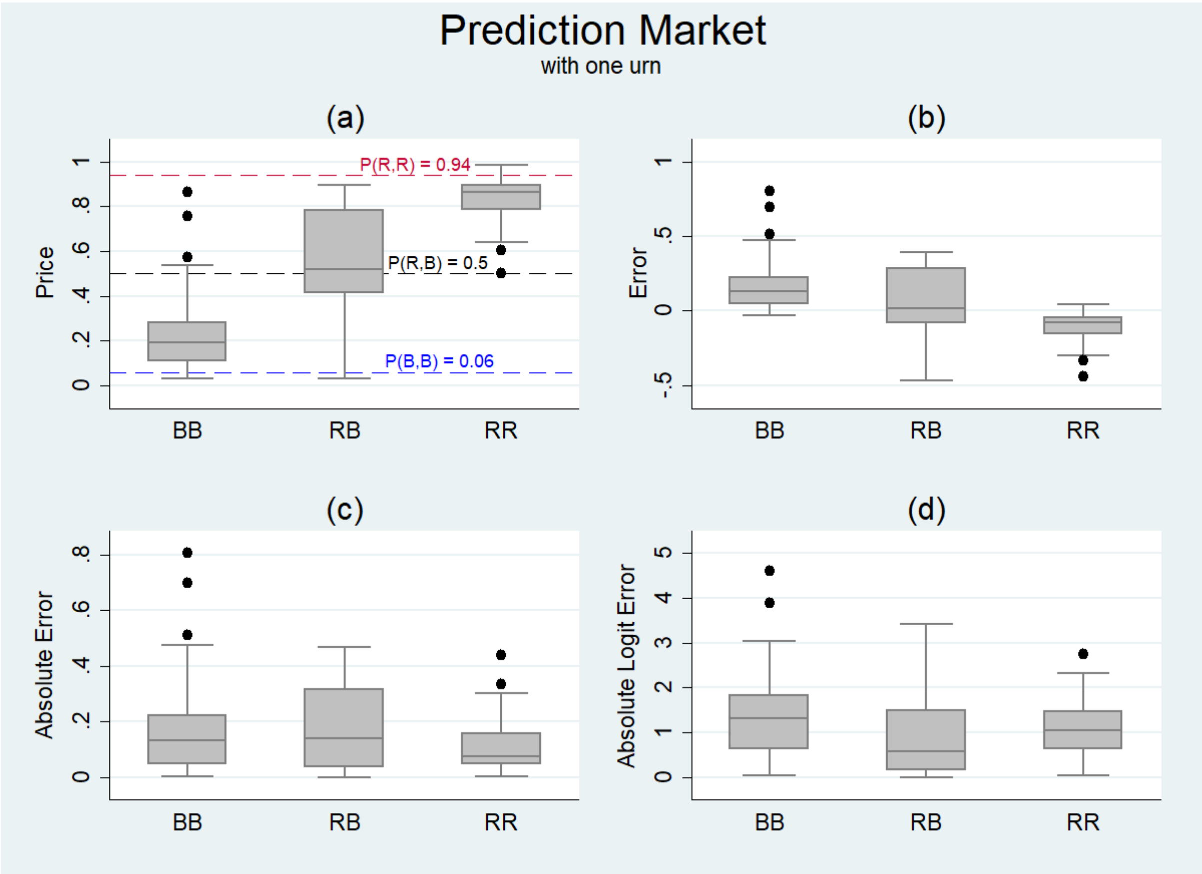


Figure 3.2 Boxplots showing the prices and errors (y-axis) depending on different signals (x-axis) provided to the participants for all markets in Stage 1. Prices shown are for assets that pay $1 for Red Type urns. The Blue Type assets are equivalent to short positions on Red Type assets, and the market maker ensures that the prices always add up to one. Signals provided are BB= Blue, Blue, RB= Red, Blue and RR = Red, Red. Panel (a) Price by Signals (b) Error by Signals (c) Absolute Error by Signals (d) Absolute Logit Error by Signals.

Figure 3.2 (a) shows that final prices depend on the signals. For two blue signals, they are distributed around 20%, for two red signals around 80%, and for one red and one blue signal they are around 50%. The variance of prices is smallest when traders receive two red signals and tends to be larger when they receive one red and one blue signal. When participants receive the same signals (RR or BB), the final prices are not driven all the way to the correctly aggregated probabilities of 94% and 6%, respectively. As shown in panel (b), the error-i.e., the difference between final market price and the correctly aggregated probability, tends to be positive for two blue signals and negative for two red signals. For one red and one blue signals, the difference is centered around zero. This demonstrates that for signals BB, the prices overestimate the correct probability whilst for signals RR, prices underestimate the correct probability.

Figure 3.2 (c) shows that the absolute error is lowest when participants receive two red signals. Participants' predictions were thus relatively close to the true value of the given signal. For signals BB and RB, the median is higher. In panel (d), when the difference between log odds is used to quantify accuracy, errors are smaller for one red and one blue signal, which is different from what we observed in (c). This illustrates that findings can differ depending on how errors are defined and underlines the importance of using alternative methods to quantify errors to test which findings are robust. The results shown in Figure 3.2. are further analysed with OLS models described in the following sections.

3.2.1 Market prices and their dependence on participants' signals

Table 3.3 displays the results of our statistical analysis of the factors influencing the final market price.

Table 3.3 Determinants of final market prices

A set of OLS models are implemented with price as our dependent variable on 240 markets. Independent variables are the treatments and signal constellations. The treatments are deterministic decision markets (DMD), stochastic decision markets (DMS) and two similitaneous prediction markets (PM2). The signal constellations are two blue signals, two red signals and one blue and one red signal. We include a dummy variable for DMD and DMS with the control treatment PM2 as the base category to demonstrate that performance is not different across the three treatments. BB is also used as a baseline across all our models. From Model (3), we also test the difference between RR and RB, and we find that it is statistically significant at 0.310, p<0.001.

Model	(1)	(2)	(3)	(4)	(5)
RB	0.322***		0.322***	0.322***	0.321***
	(0.0266)		(0.0267)	(0.0266)	(0.0277)
RR	0.632***		0.632***	0.632***	0.627***
	(0.0266)		(0.0267)	(0.0266)	(0.0391)
DMD		0.0199	0.0199	0.0189	0.0199
		(0.0476)	(0.0253)	(0.0253)	(0.0253)
DMS		0.00415	0.00415	0.00230	0.00415
		(0.0476)	(0.0253)	(0.0253)	(0.0253)
SessionID				0.00370	
				(0.00299)	
Round					-0.00204
					(0.0107)
Constant	0.204***	0.387***	0.196***	0.173***	0.204***
	(0.0133)	(0.0336)	(0.0198)	(0.0272)	(0.0440)
Observations	240	240	240	240	240
R^2	0.719	0.001	0.720	0.722	0.720
Adj. R^2	0.717	-0.008	0.715	0.716	0.714

Notes: Standard errors in parentheses. Significance at *.05, **.01, ***.001.

Result 1. Final market prices are dependent on private signals.

The results show that market prices clearly depend on the private signals. Compared to the prices for two blue signals, prices are statistically significantly higher for the other signal constellations (0.322 for RB and 0.632 for RR, $p<0.001$). This dependence is not affected by potential effects from rounds, Session ID and the treatments used in Stage 2, and is in line with what is observed in Figure 3.2. We find no statistically significant impact of Stage 2 treatments, SessionID or Round on final market prices once we introduce them into our models (Model 2- Model 5).

When all traders in the market receive the same private signals as opposed to conflicting signals (RR and BB vs. RB), it is highly likely that the information driving the market matches all traders' private signals. As a result, the final market prices of the same private signals would be more significantly impacted.

3.2.2 Quantifying information aggregation through forecasting error

Absolute error

Table 3.4 displays the results of our analysis of the factors influencing our absolute error. We test whether there are significant effects across our different treatments. Since Stage 1 is used as a control and no decision rule is employed (all treatments are the same), we expect no treatment effects.

Table 3.4 Determinants of absolute error

A set of OLS models are implemented with absolute error as our dependent variable on 240 markets. We include a dummy variable for DMD and DMS with the control treatment PM2 as the base category to demonstrate that performance is not different across the three treatments. BB is also used as a baseline across all our models. From Model (3), we also test the difference between RR and RB, and we find that it is statistically significant at 0.0881, p<0.01.

Model	(1)	(2)	(3)	(4)	(5)
RB	0.0505*		0.0505*	0.0505*	0.0486*
	(0.0219)		(0.0218)	(0.0217)	(0.0226)
RR	-0.0376		-0.0376	-0.0376	-0.0451
	(0.0219)		(0.0218)	(0.0217)	(0.0319)
DMD		-0.00225	-0.00225	-0.00312	-0.00225
		(0.0211)	(0.0207)	(0.0206)	(0.0207)
DMS		0.0337	0.0337	0.0319	0.0337
		(0.0211)	(0.0207)	(0.0207)	(0.0207)
SsessionID				0.00349	
				(0.00244)	
Round					-0.00281
					(0.00875)
Constant	0.146***	0.138***	0.136***	0.114***	0.146***
	(0.0109)	(0.0149)	(0.0162)	(0.0222)	(0.0359)
Observations	240	240	240	240	240
R^2	0.044	0.015	0.059	0.068	0.060
Adj. R^2	0.036	0.007	0.043	0.048	0.040

Notes: Standard errors in parentheses. Significance at *.05, **.01, ***.001. PM2 and BB are used as baselines across all our models.

We find a statistically significant relationship between the error and combined signals for the signal constellations of RB. Compared to the error of the two blue signals, error is statistically significantly higher for RB (0.0505, p<0.05) and statistically insignificant for RR (-0.0376). The difference between RR and RB is also statistically significant and this demonstrates that prediction performance is better when participants receive two red balls or two blue balls rather than one red and one blue ball. This could also be an indication that participants are more confident about their investment decisions when they receive same signal constellations rather than conflicting ones.

Although Stage 2 occurred after Stage 1 and so cannot have any causal influence over Stage 1 results, we nevertheless looked for differences in Stage 1 performance according to Stage 2 settings (the type of decision rule). This was to establish whether any systematic differences between Stage 2 results were due to intrinsic differences between the groups. We tested this (see models 2-5) to ensure that there is no difference as this will affect interpretation of results on treatment effects in the decision market. Our expectation that there would be no such difference holds since we find that the effects of Stage 2 treatments are non-significant from Stage 1 errors, which indicates that skills amongst participants are evenly distributed across the Stage 2 treatments. This provides a good control to determine whether any observed differences in Stage 2 are real differences. We additionally included SessionID (model 4) and Round (model 5) and observe no effect from these variables.

Absolute logit error

Table 3.5 displays the results of our robustness analysis using the absolute error of logit transformed probabilities. We take the difference of the log odds of the final market prices and the correct Bayesian probability instead of taking the differences between probabilities. This transforming increases errors for probabilities that are closer to zero or one and makes the distribution of errors more similar to a normal distribution. Amplifying the error for extreme probabilities demonstrates that using absolute error may not be the best or only method to determine error. From our second hypothesis and from the results from Table 3.4, we check whether treatment effects are still absent and if a positive relationship between error and private signals still exists.

Table 3.5 Robustness of results using an alternative error measure

A set of OLS models are implemented with absolute error of logit transformed probabilities as our dependent variable on 240 markets. We include a dummy variable for DMD and DMS with the control treatment PM2 as the base category to demonstrate that performance is not different across the three treatments. BB is also used as a baseline across all our models. From Model (3), we also test the difference between RR and RB, and we find that it is not statistically significant at 0.102.

Model	(1)	(2)	(3)	(4)	(5)
RB	-0.272*		-0.272*	-0.272*	-0.284*
	(0.129)		(0.128)	(0.128)	(0.133)
RR	-0.170		-0.170	-0.170	-0.216
	(0.129)		(0.128)	(0.128)	(0.188)
DMD		0.00450	0.00450	0.0000656	0.00450
		(0.123)	(0.122)	(0.122)	(0.122)
DMS		0.221	0.221	0.212	0.221
		(0.123)	(0.122)	(0.122)	(0.122)
SessionID				0.0177	
				(0.0144)	
Round					-0.0175
					(0.0515)
Constant	1.248***	1.085***	1.173***	1.062***	1.238***
	(0.0645)	(0.0867)	(0.0952)	(0.131)	(0.212)
Observations	240	240	240	240	240
R^2	0.021	0.018	0.039	0.045	0.039
Adj. R^2	0.013	0.009	0.022	0.024	0.019

Notes: Standard errors in parentheses. Significance at *.05, **.01, ***.001. PM2 and BB are used as baselines across all our models.

We observe that the error gets smaller for signals RB, compared to BB (i.e., the effect is negative rather than positive) which is different from what we previously observed in Table 3.4. Stage 2 treatments still have no statistically significant impact on accuracy. This means our findings for an approximately equal distribution of participants skills is not affected by the way accuracy is quantified. We notice a small (non-significant) effect in the stochastic setting where in comparison to the results from Table 3.4, it had a slightly larger error. This could signify that those participants in the stochastic setting are to some extent less skillful than participants in other settings. Furthermore, we still find no statistically significant effect across the models that introduce SessionID and Round. Overall, regardless of how error is measured, our findings are not affected; there is still no correlation between market error and Stage 2 treatments but there is a relation between market error and signal constellations, reconfirming our Result 1 that market prices depend on the private signals of participants.

3.3 Results from Stage 2

In Stage 2, participants traded in a decision market with independently drawn signals from two urns. Each group consisted of four participants. Two participants received independently drawn signals from urn 1 and the two other participants received independently drawn signals from urn 2.

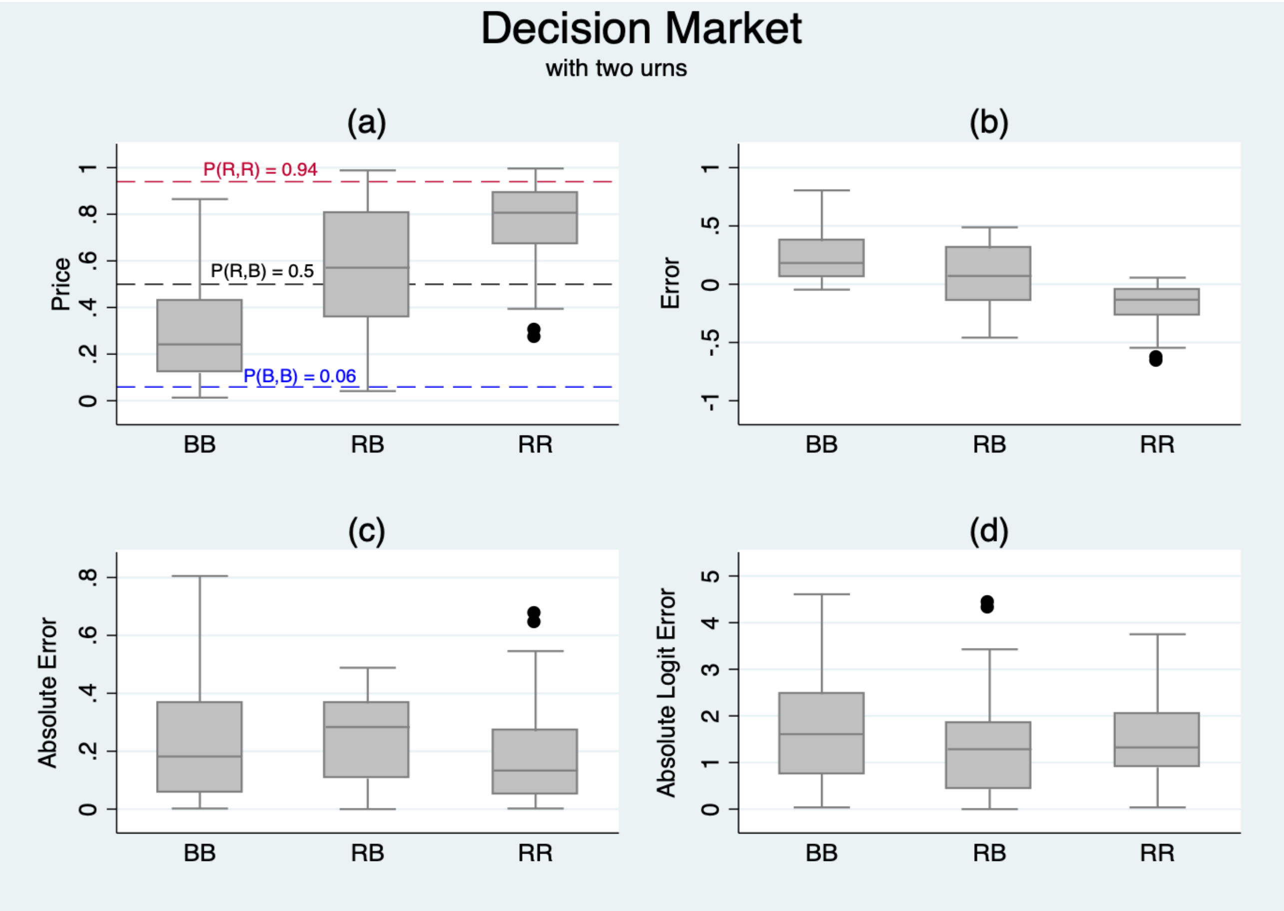


Figure 3.3 Boxplots showing the prices and errors (y-axis) depending on different signals (x-axis) provided to the participants for all treatments in Stage 2. Prices shown are for assets that pay $1 for Red Type urns. The Blue Type assets are equivalent to short positions on Red Type assets, and the market maker ensures that the prices always add up to one. Signals provided are BB= Blue, Blue, RB= Red, Blue and RR = Red, Red. Panel (a) Price by Signals (b) Error by Signals (c) Absolute Error by Signals (d) Absolute Logit Error by Signals.

Figure 3.3 (a) shows that similarly to the markets in Stage 1, the final prices in Stage 2 markets depend on the signals. For two blue signals, they are distributed around 25%, for two red signals around 80%, and for one red and one blue signal they are around 60%. In comparison to variances in Stage 1, depicted in Figure 3.2 (a), the variance of prices here is slightly larger when traders receive same signal constellations (RR or BB). The variance remains largest when participants receive one red and one blue signal. As in Stage 1, when participants receive two blue signals, prices are overestimated and when signals are RR, the final prices are underestimated (the final prices are not driven all the way to the correctly aggregated probability of 94%, in fact, they are underestimated). As shown in panel (b), the error tends to be positive for two blue signals, slightly positive for one red and one blue signals, and negative for two red signals. Compared to Figure 3.3 (b), we observe that the error is not centered around zero but is slightly higher, making it positive.

Furthermore, in Figure 3.3 (c), the absolute error is lowest when participants receive two red signals. For signals RB, the median error is highest. In panel (d), when the difference between log odds is used, errors are smaller for one red and one blue signal, which is different for what we observed in (c) but similar to what we observed from our Stage 1. This illustrates the robustness of using log odds as an alternative method to quantify error.

3.3.1 Market prices and their dependence on participants' signals

Table 3.6 displays the results of our statistical analysis of the factors influencing the final market price.

Table 3.6 Final market prices and their dependence on signals

A set of OLS models are implemented with price as our dependent variable on 384 markets. Independent variables are the treatments and signal constellations. The treatments are deterministic decision markets (DMD), stochastic decision markets (DMS) and two similitaneous prediction markets (PM2). The signal constellations are two blue signals, two red signals and one blue and one red signal. We include a dummy variable for DMD and DMS with the control treatment PM2 as the base category to demonstrate that performance is not different across the three treatments. BB is also used as a baseline across all our models. From Model (3), we test the difference between RR and RB, and we find that it is statistically significant at 0.204, p<0.001. The difference between DMD and DMS is statistically significant at 0.0626, p<0.05.

Model	(1)	(2)	(3)	(4)	(5)	(6)
RB	0.284***		0.284***	0.284***	0.264***	0.283***
	(0.0279)		(0.0275)	(0.0275)	(0.0311)	(0.0307)
RR	0.489***		0.489***	0.489***	0.511***	0.488***
	(0.0320)		(0.0315)	(0.0315)	(0.0355)	(0.0326)
DMD		0.0978**	0.0978***	0.0991***	0.0978***	0.0978***
		(0.0369)	(0.0288)	(0.0288)	(0.0288)	(0.0288)
DMS		0.0352	0.0352	0.0378	0.0352	0.0352
		(0.0369)	(0.0288)	(0.0288)	(0.0288)	(0.0288)
SessionID				-0.00521		
				(0.00341)		
Round					-0.0102	
					(0.00746)	
Urn						0.00155
						(0.0262)
Constant	0.279***	0.481***	0.234***	0.267***	0.283***	0.232***
	(0.0213)	(0.0261)	(0.0268)	(0.0342)	(0.0447)	(0.0413)
Observations	384	384	384	384	384	384
R^2	0.388	0.019	0.407	0.410	0.410	0.407
Adj. R^2	0.385	0.013	0.400	0.402	0.402	0.399

Notes: Standard errors in parentheses. Significance at *.05, **.01, ***.001.

The results show that market prices strongly depend on the private signals. Compared to the prices for two blue signals, prices are statistically significantly higher for the other signal constellations (0.284 for RB and 0.489 for RR, P < 0.001). This dependence is not affected by potential effects from rounds, Session ID or the number of urns, and is consistent from what we observed in Figure 3.3. We find a statistically significant impact of our deterministic treatment (DMD) on prices once we start introducing it in our second model.

3.3.2 Quantifying information aggregation by decision rule

Absolute error

Table 3.7 displays the results of our analysis of the factors influencing the absolute error. We test whether there are significant effects across our different treatments. We expect to find treatment effects depending on the different decision rules.

Table 3.7 Market error and its dependence on the decision rule

A set of OLS models are implemented with absolute error as our dependent variable on 384 markets. We include a dummy variable for DMD and DMS with the control treatment PM2 as the base category. BB is also used as a baseline across all our models. From Model (3), we test the difference between RR and RB, and we find that it is statistically significant at -0.0659, p<0.01. This is similar to what we find in Stage 1 results (Table 3.4). We find no statistically significant effect between DMD and DMS (0.0282).

Model	(1)	(2)	(3)	(4)	(5)	(6)
RB	0.0259		0.0259	0.0259	0.0128	0.0397
	(0.0198)		(0.0197)	(0.0197)	(0.0222)	(0.0219)
RR	-0.0401		-0.0401	-0.0401	-0.0253	-0.0320
	(0.0227)		(0.0225)	(0.0226)	(0.0254)	(0.0232)
DMD		0.0539**	0.0539**	0.0540**	0.0539**	0.0539**
		(0.0208)	(0.0206)	(0.0206)	(0.0205)	(0.0205)
DMS		0.0258	0.0258	0.0260	0.0258	0.0258
		(0.0208)	(0.0206)	(0.0206)	(0.0205)	(0.0205)
SessionID				-0.000439		
				(0.00244)		
Round					-0.00673	
					(0.00533)	
Urn						-0.0269
						(0.0187)
Constant	0.223***	0.198***	0.196***	0.199***	0.229***	0.229***
	(0.0151)	(0.0147)	(0.0191)	(0.0245)	(0.0319)	(0.0294)
Observations	384	384	384	384	384	384
R^2	0.025	0.017	0.042	0.042	0.046	0.047
Adj. R^2	0.020	0.012	0.032	0.030	0.034	0.035

Notes: Standard errors in parentheses. Significance at *.05, **.01, ***.001.

Result 2. Market error depends on the decision rule used.

We find no statistically significant relationship between the error and the signal constellations. The errors are higher compared to our results from Stage 1. In model 1 from Table 3.4, our constant estimate is 0.146, whilst in the same model from Table 3.7 our constant is 0.223. This illustrates that market error estimates tend to be larger in our more complex Stage 2 treatments.

We observe that the market error depends significantly on the decision rules; compared to the error in our PM2 setting, error is statistically significantly higher for the deterministic decision rule (0.0539, $p < 0.01$). This dependence is not affected by potential effects from rounds, SessionID or the number of urns.

Absolute logit error

Table 3.8 displays the results of our robustness analysis using the absolute error of logit transformed probabilities. From the results from Table 3.7, we check whether treatment effects still hold.

Table 3.8 Robustness of treatment effects using an alternative error measure

A set of OLS models are implemented with absolute error of logit transformed probabilities as our dependent variable on 384 markets. We include a dummy variable for DMD and DMS with the control treatment PM2 as the base category. BB is also used as a baseline across all our models. From Model (3), we test the difference between RR and RB, and we find no statistical significance (0.226). We find no statistically significant difference between DMD and DMS (-0.144).

Model	(1)	(2)	(3)	(4)	(5)	(6)
RB	-0.350**		-0.350**	-0.350**	-0.459***	-0.285*
	(0.114)		(0.113)	(0.113)	(0.128)	(0.126)
RR	-0.124		-0.124	-0.124	-0.000338	-0.0856
	(0.131)		(0.130)	(0.130)	(0.146)	(0.134)
DMD		0.309*	0.309**	0.310**	0.309**	0.309**
		(0.120)	(0.118)	(0.119)	(0.118)	(0.118)
DMS		0.164	0.164	0.166	0.164	0.164
		(0.120)	(0.118)	(0.119)	(0.118)	(0.118)
SessionID				-0.00425		
				(0.0140)		
Round					-0.0561	
					(0.0306)	
Urn						-0.127
						(0.108)
Constant	1.638***	1.296***	1.480***	1.506***	1.749***	1.633***
	(0.0870)	(0.0846)	(0.110)	(0.141)	(0.183)	(0.170)
Observations	384	384	384	384	384	384
R^2	0.025	0.017	0.043	0.043	0.051	0.046
Adj. R^2	0.020	0.012	0.033	0.030	0.039	0.034

Notes: Standard errors in parentheses. Significance at *.05, **.01, ***.001. PM2 and BB are used as baselines across all our models.

Once we alter our error definition, we still observe the same effects, but when signal constellations are RB, the market error affects our model differently. There is a significant effect when signals are RB; the error becomes smaller compared to BB, i.e., switches from positive (in Table 3.7) to negative. Once we factor signal constellations in our models, we notice that our deterministic setting remains statistically significant. This is in line with our results from Table 3.7. Overall, this shows that our findings on how the decision rule affects the market performance do not depend on how we define our error. Market error remains lower for our stochastic setting and the PM2 setting than for our deterministic one. We still find no statistically significant effect across the models that introduce Session ID, round, or urns.

3.3.3 Prices and errors under different settings and different signal constellations

To further investigate the differences in error and prices in the different decision market settings, we extended the analyses presented in Section 3.3.2 by analysing the dependence of prices and errors on the settings separately for the different signal constellations.

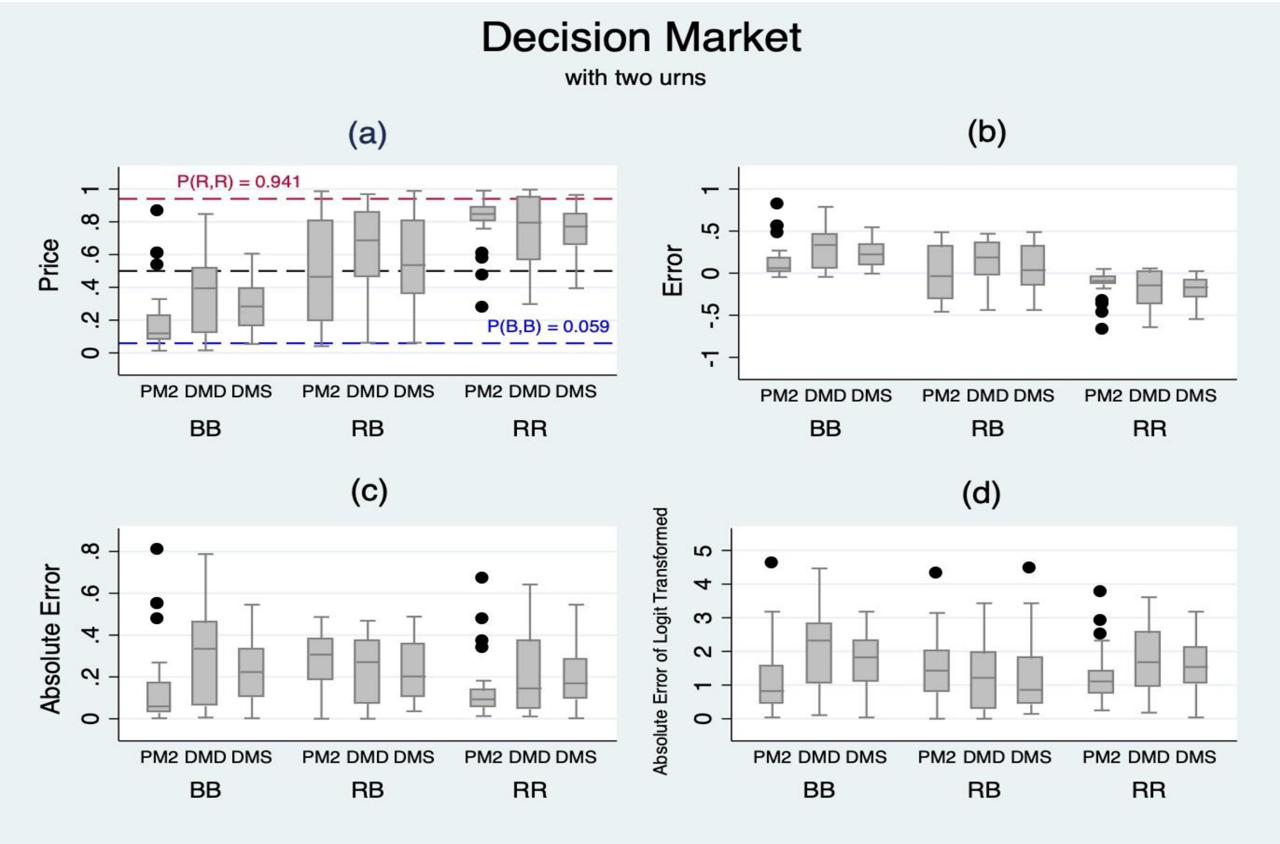


Figure 3.4 Boxplots showing the prices and errors (y-axis) depending on different settings and different signals (x-axis) provided to the participants for all treatments in Stage 2. Signals provided are BB= Blue, Blue, RB= Red, Blue and RR = Red, Red. Panel (a) Price by Setting and Signals (b) Error by Setting and Signals (c) Absolute Error by Setting, and Signals (d) Absolute Error of Logit Transformed by Setting and Signals.

Figure 3.4 (a) shows that the final prices in Stage 2 settings depend on the signals that participants receive. Final prices are higher for Setting DMD, specifically for signals BB. Under DMD, for two blue signals, final prices are distributed around 40%, 70% for one red and one blue signal and around 80% for two red signals.

This observation is also depicted in 3.4 (c) and (d) where absolute error and absolute error of logit transformed are highest for the DMD setting when participants receive two blue signals. Participants appear to avoid purchasing Blue Type assets when they receive two blue signals.

Table 3.9 displays the results of our statistical analysis of the dependence of prices and errors under different settings when participants receive two blue signal constellations.

Table 3.9 Prices and errors under different settings for two blue signals

A set of OLS models are implemented with price and errors as our dependent variables on the 120 markets that participants receive two blue signals. We include a dummy variable for DMD and DMS with the control treatment PM2 as the base category to understand participant's decisions across the three treatments. From Model (1), we test the difference between DMD and DMS and we find no statistically significant effect (-0.0762). From Model (2), the difference between DMD and DMS is statistically significant at -0.0824, p<0.05 and from Model (3), the difference is -0.182.

Dependent Variable	Model(1)	Model(2)	Model(3)
	Price	AbsoluteError	AbsoluteLogitError
DMD	0.179***	0.179***	0.946***
	(0.0418)	(0.0406)	(0.220)
DMS	0.102*	0.0968*	0.609**
	(0.0418)	(0.0406)	(0.220)
Constant	0.185***	0.131***	1.119***
	(0.0296)	(0.0287)	(0.156)
Observations	120	120	120
R^2	0.136	0.143	0.139
Adj.R^2	0.121	0.128	0.125

Notes: Standard errors in parentheses. Significance at *.05, **.01, ***.001. PM2 is used as a baseline across all our models.

The results from Model (1) show that prices for markets with two blue signals are substantially higher for our deterministic treatment compared to our PM2 setting. We find a statiscally significant effect at 0.179, p<0.001, and this is consistent with what we observed in Figure 3.4 (a). Results from Model (1) translate into the large errors we observe in Model (2). We find a statistically significant relationship between the absolute error in DMD for BB at 0.179, p<0.001 compared to our PM2 setting. Even when we alter our error definition in Model (3), we still observe the same robust significant effects (0.946, p<0.001). This is also illustrated in Figure 3.3 (b) and (c), respectively. Clearly participants prefer holding Red Type assets even though their private signals could be blue. This could largely be due to them observing the behavior of other participants in the market who trade on Red Type assets (increasing the price of Red Type assets and decreasing those of the blue). We also test the effects for signal constellations RB and RR in Tables 3.10 and 3.11 below.

Table 3.10 Prices and errors under different settings for RB

A set of OLS models are implemented with price and errors as our dependent variables on the 168 markets that participants receive one blue and one red signal. We include a dummy variable for DMD and DMS with the control treatment PM2 as the base category.

Dependent Variable	Model(1)	Model(2)	Model(3)
	Price	AbsoluteError	AbsoluteLogitError
DMD	0.133*	-0.0487	-0.265
	(0.0521)	(0.0264)	(0.173)
DMS	0.0531	-0.0533*	-0.281
	(0.0521)	(0.0264)	(0.173)
Constant	0.501***	0.283***	1.469***
	(0.0369)	(0.0187)	(0.122)
Observations	168	168	168
R^2	0.039	0.029	0.020
Adj. R^2	0.027	0.018	0.008

Notes: Standard errors in parentheses. Significance at *.05, **.01, ***.001.
PM2 is used as a baseline across all our models.

Compared to our results from Table 3.9, the results for signal constellations RB are less significant. From Model (1), prices remain higher for our DMD setting compared to PM2 with a significant effect of 0.133, $p<0.05$. Despite this significant effect, this does not translate into larger errors (in Models (2) and (3)). Errors move in the same direction and are not larger compared to our PM2 setting. Meanwhile, under our stochastic market settings, the results are similar to PM2 in terms of pricing (Model (1)) and the absolute error is a bit smaller (Model (2)). However, despite the significant effect we observe from DMS in Model (2) at -0.0533, $p<0.05$, this effect is relatively small and disappears once we start using our logit error measure in Model (3).

Table 3.11 Prices and errors under different settings for two red signals

A set of OLS models are implemented with price and errors as our dependent variables on the 96 markets that participants receive two red signals. We include a dummy variable for DMD and DMS with the control treatment PM2 as the base category.

Dependent Variable	Model(1)	Model(2)	Model(3)
	Price	AbsoluteError	AbsoluteLogitError
DMD	-0.0650	0.0768	0.515*
	(0.0437)	(0.0409)	(0.214)
DMS	-0.0800	0.0753	0.386
	(0.0437)	(0.0409)	(0.214)
Constant	0.816***	0.132***	1.213***
	(0.0309)	(0.0289)	(0.151)
Observations	96	96	96
R^2	0.039	0.047	0.063
Adj. R^2	0.018	0.027	0.043

Notes: Standard errors in parentheses. Significance at *.05, **.01, ***.001. PM2 is used as a baseline across all our models.

From Model (1), prices are somewhat lower for our DMD setting under the RR signal constellations, but the absolute error is a bit higher but with a relatively small effect (Model (2)). We find a significant effect from using our logit error measure in Model (3) at 0.515, $p<0.05$ for the DMD setting compared to the PM2 setting. Overall, our results from Tables 3.10 and 3.11 are less statistically significant compared to our results from Table 3.9. Effects for signal constellations BB are statistically significant compared to the other signal constellations.

3.4 Discussions and Conclusion

In Stage 1, our main result (Result 1) is that market prices are dependent on private signals. This signifies that the signal constellation is the main determinant for the final prices in our markets. Whilst we have systematic errors and some deviations from the correct Bayesian aggregation, it is clear that final market prices for the different signal constellations are different. This illustrates that our market clearly incorporates the ball signals. We observe the same in Stage 2, where our more complex markets have larger errors, but result in prices that are closely related by the signals.

Our second important finding (Result 2) is that in Stage 2, errors depend on the decision rule. In comparison to our setting with two simultaneous prediction markets (PM2), the error is significantly larger for our deterministic setting (DMD). For the stochastic decision rule (DMS), the error lays in the middle between our PM2 and DMD setting. The errors in the stochastic setting is not significantly different from the prediction market setting, and is also not significantly different from the deterministic setting. The observation that the errors are highest for the deterministic decision rule indicates that incentive compatibility of the decision rule matters for information aggregation in our experimental markets. The observation that errors for the stochastic decision rule tend to be larger compared to the PM2 setting suggests that the higher complexity of the decision rule might be associated with the large error.

Our models in Table 3.3 (Stage 1) in comparison to Table 3.6 (Stage 2) shows that not only error, but also prices are higher for our deterministic setting. This suggests that the increased error could be driven by the inflated prices in the settings, in particular when the signals are BB or RB. Moreover, there is no effect of the Stage 2 decision rule on the error in Stage 1, signifying that participants skills tend to be evenly distributed across the settings and does not explain the differences observed in Stage 2. Furthermore, market error depends on the signal constellations, but this dependence is not robust to the definition of error. When assessed through the absolute error, the markets are more accurate when the participants receive the same (RR, BB) rather than different signals (RB). However, when assessed through the absolute error of logit transformed probabilities, the opposite is observed because the logit transformation magnifies errors for more extreme probabilities. This illustrates why it is important to include a second error measure as a robustness test.

Following up on Result 2, we further explored how final prices and errors in the Stage 2 settings depend on the signals that participants receive. The results of our analysis in Table 3.9 show that prices tend to be higher in DMD compared to PM2, in particular for signal constellations BB, where it translated into a substantially increased error. This suggests that possibly, in the deterministic markets, participants avoid holding contracts for Blue Type urns. This could be because purchasing such contracts makes it less likely that the market will resolve. This will be tested in an additional analysis in Chapter 4.

Chapter 4: Individual Behavior in the Prediction Markets Experiment: The Relation to Participant Characteristics and Market Accuracy

"Economics, when you strip away the guff and mathematical sophistry, is largely about incentives." — John Cassidy, How Markets Fail: The Logic of Economic Calamities

4.1 Introduction

In this chapter we describe the second part of our experimental results from the prediction market experiment as outlined in Chapter 2. We investigate if there are behavioral attributes associated with trading performance by addressing the following three hypothesis:

Hypothesis 1: Is there evidence of individual differences in trading behavior?
Hypothesis 2: Are individual differences in trading behavior related to individual attributes as measured in Stage 3?
Hypothesis 3: Do these individual differences influence market accuracy?

For our analysis, we use the Stage 1 prediction market data, because unlike for Stage 2, all 96 participants participated in the same setting. Since in all our experiments, we have for each individual trade recorded the traded assets, execution time and purchasing price, we can construct from this record parameters that capture the behavior of a trader in a market. This approach provides us with a considerable number of observations to link trading behavior with individual characteristics and with market performance. In the following, we outline how the three hypotheses above are investigated in the subsequent sections, and describe the variables used in our analyses.

In section 4.2 we investigate the first of the three hypotheses by implementing factorial ANOVAs for a set of ordinary least-square (OLS) models with four different individual trading parameters as dependent variables (see Table 4.1) and test if there are individual differences in trading behavior. We use the number of trades a participant executed in a market to quantify how actively the participant is trading. Since the timing of a trade likely impacts pricing and thus the costs and benefits of a trade, we additionally use the number of trades in the first twenty seconds and the time of the first trade executed in a market. We use a time period of twenty seconds because we observe that trading has a peak intensity around 10-15 seconds (see figure 4.1). Half of the trades occur in the first twenty seconds, which is almost twice the rate of trading as in the second half (from 20 seconds onwards), and therefore we use the number of trades in the first twenty seconds as an indicator for how much participants engaged in early trading. Furthermore, to capture what assets a participant purchased, we analyze the final portfolio holdings and calculate the expected value of the holdings for each participant in each market, using the correctly aggregated probability given both signals. The results of the ANOVAs are summarized in Table 4.2, and further illustrated in Figure 4.2.

In section 4.3 we address our second hypothesis by exploring the experimental evidence on risk aversion and demographic characteristics of participants that could explain the variation in trading behavior. We use the individual averages of the four primary trade parameters used in section 4.2 over the five Stage 1 markets a participant traded in, and calculated the average number of trades per participant, average number of trades in the first twenty seconds, average time of the first trade and the average expected value of final portfolio holdings (see Table 4.4). To assess risk aversion, we use different analysis measures of the Multiple Price List (MPL) by Holt and Laury (2002). As described in Chapter 2, risk attitudes are elicited in Stage 3 through adopting the MPL task, where an ordered array of 10 binary lottery choices is presented to participants. We use two proxies, the first adopted from Holt and Laury (2002) and the second from Eckel et al. (2010): the number of safe choices, and inconsistency in the choices. The number of safe choices describes how often a participant selects option A (the safe option) rather than option B (the risky option) in the paired lottery tasks. The more safe choices a participant makes, the more risk averse they are considered to be. Inconsistency indicates if a participant has more than one switching point in the MPL task (see Appendix G for a more elaborative description.) As results from the authors suggest that different elicitation methods may produce different results, we chose two proxies instead of one to provide a more robust understanding of the data. Participant characteristics are used to complement our understanding of risk attitudes whilst allowing further analysis of behavior in our prediction market experiment. A description of the variables used in this section is given in Table 4.4. The results from the statistical models are summarized in Table 4.5 and 4.6.

The results for the third hypothesis, i.e., whether individual differences in trading behavior influence market accuracy, are presented in Section 4.4. We again employ variations of the four variables from Sections 4.2 and 4.3, but instead of looking at individual averages, we aggregate the variables on a market level. These variables are the total number of trades of the two participants in a market, the total number of trades for both participants in the first twenty seconds, the time of the first trade executed by a participant in a market, and the total expected value of final portfolio holdings for the two participants in a market. We quantify market efficiency by the absolute error and the absolute error of logit transformed probabilities (see Chapter 3 Table 3.1 for the definition of errors).

In section 4.5, we use in contrast to the other analyses presented in this chapter data from the Stage 2 markets. We follow up on the findings from Chapter 3 Section 3.8 about errors in the DMD setting in the Stage 2 markets by investigating how the participants' behavior differs between the different settings in this stage. Finally, we conclude and discuss the findings of this chapter in section 4.6.

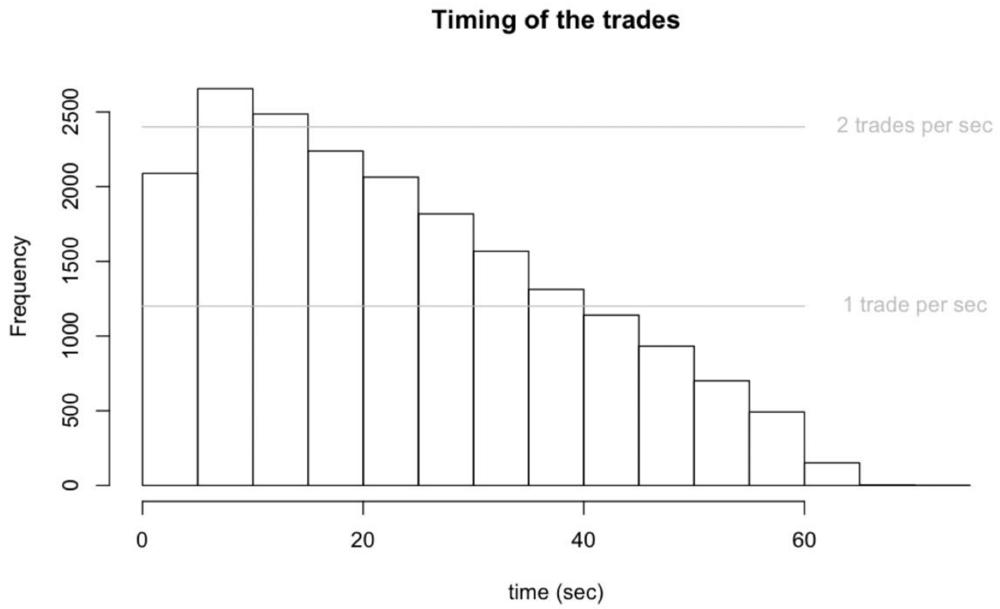


Figure 4.1 Timing of the approximately 20,000 trades in our Stage 1 prediction markets. Each bar represents a 5 second period that covers 240 markets. Grey lines indicate the number of trades scaled down to a single market.

4.2 Individual Variation in Trading Behavior

Factorial ANOVAs were conducted for a sample of 480 observations from 96 participants to examine the individual trading behavior of the participants. Four separate ANOVAs were conducted for the four dependent variables: number of trades per participant (NumTrades), number of trades in the first twenty seconds (NumTrades20), time of the first trade (FirstTrade) and the expected value of final portfolio holdings (PortfolioValue). Each ANOVA uses three independent variables: SessionID, Round (see Table 3.1 in Chapter 3) and PartcipantID (see Table 4.1).

Table 4.1 Description of variables for assessing individual variation in trading behaviour

Variable name	Description
NumTrades	The number of trades per participant per market
NumTrades20	The number of trades in the first twenty seconds
FirstTrade	The time of the first trade executed in a market
PortfolioValue	Expected value of final portfolio holdings for each participant in a market
ParticipantID	A generated identification number to each participant

ParticipantID is entered into the ANOVAs as a factor to identify how much individual variation exists between different participants. We include the number of rounds as an independent variable since performance and behavior could change over the course of the experimental stage. We also include SessionID to observe any experimenter effects that might have caused an impact on trading over the duration of the entire experiment. These experimenter effects might include a difference in speed or explanation of the presented instructions.

Table 4.2 Summary of ANOVA results on trading performance

	Model (1)					Model (2)				
Dependent Variable	NumTrades					NumTrades20				
	Seq. SS	*df*	*MS*	*F*	*p-value*	*Seq. SS*	*df*	*MS*	*F*	*p-value*
Source										
SessionID	19.15	1	19.15	0.52	0.4695	977.83	1	977.83	76.31	0.0000
Round	25.35	1	25.35	0.69	0.4054	5.4	1	5.4	0.42	0.5166
ParticipantID	14613.43	94	155.46	4.25	0.0000	14132.43	94	150.34	11.73	0.0000
Error	13994.65	383	36.53			4907.8	383	12.81		
R^2	0.5116					0.7549				
Adj. R^2	0.3891					0.6935				

	Model (3)					Model (4)				
Dependent Variable	FirstTrade					PortfolioValue				
	Seq. SS	*df*	*MS*	*F*	*p-value*	*Seq. SS*	*df*	*MS*	*F*	*p-value*
Source										
SessionID	20.94	1	20.94	15.18	0.0001	7.13	1	7.13	1.32	0.2521
Round	2.50	1	2.50	1.81	0.1790	.108	1	.108	0.02	0.8874
ParticipantID	915.07	94	9.73	7.06	0.0000	791.39	94	8.41	1.55	0.0022
Error	528.26	383	1.37			2076.20	383	5.42		
R^2	0.6398					0.2778				
Adj. R^2	0.5496					0.0968				

Notes: Seq. SS = Sequential Sum of Squares, *df* = degrees of freedom, MS = Mean Squares

In Table 4.2, model (1) presents the ANOVA results for the number of trades per participant in a market, model (2) shows the results for the total number of trades in the first twenty seconds, model (3) shows the results for the time of the first trade executed in a market and model (4) shows the results with the expected value of the final portfolio of a participant in a market.

Result 1: Variation exists amongst participants' trading behavior.

Our ANOVA results show that a large fraction of variation can be attributed to between-participant variation. The results from all our models indicating that the variation of the dependent variable is explained by our three independent variables, with most of the variation being attributed to ParticipantID, indicating that trading significantly varies according to each participant. In models (1), (2) and (3) the sequential sum of squares for Participant ID is larger than the remaining error in our models. In (a), 51% of the variation in the number of trades is explained by ParticipantID, in (2) it is 75% and in (3) it is 63%. In model (4), the variation explained through participant ID in portfolio values is smaller than the variation we tend to observe in the other trading parameters (models (1) to (3)), with approximately 28% of the variation in the portfolio value being explained by ParticipantID. This indicates that there are less individual differences in the value of the final portfolio value than there are for the other trading parameters.

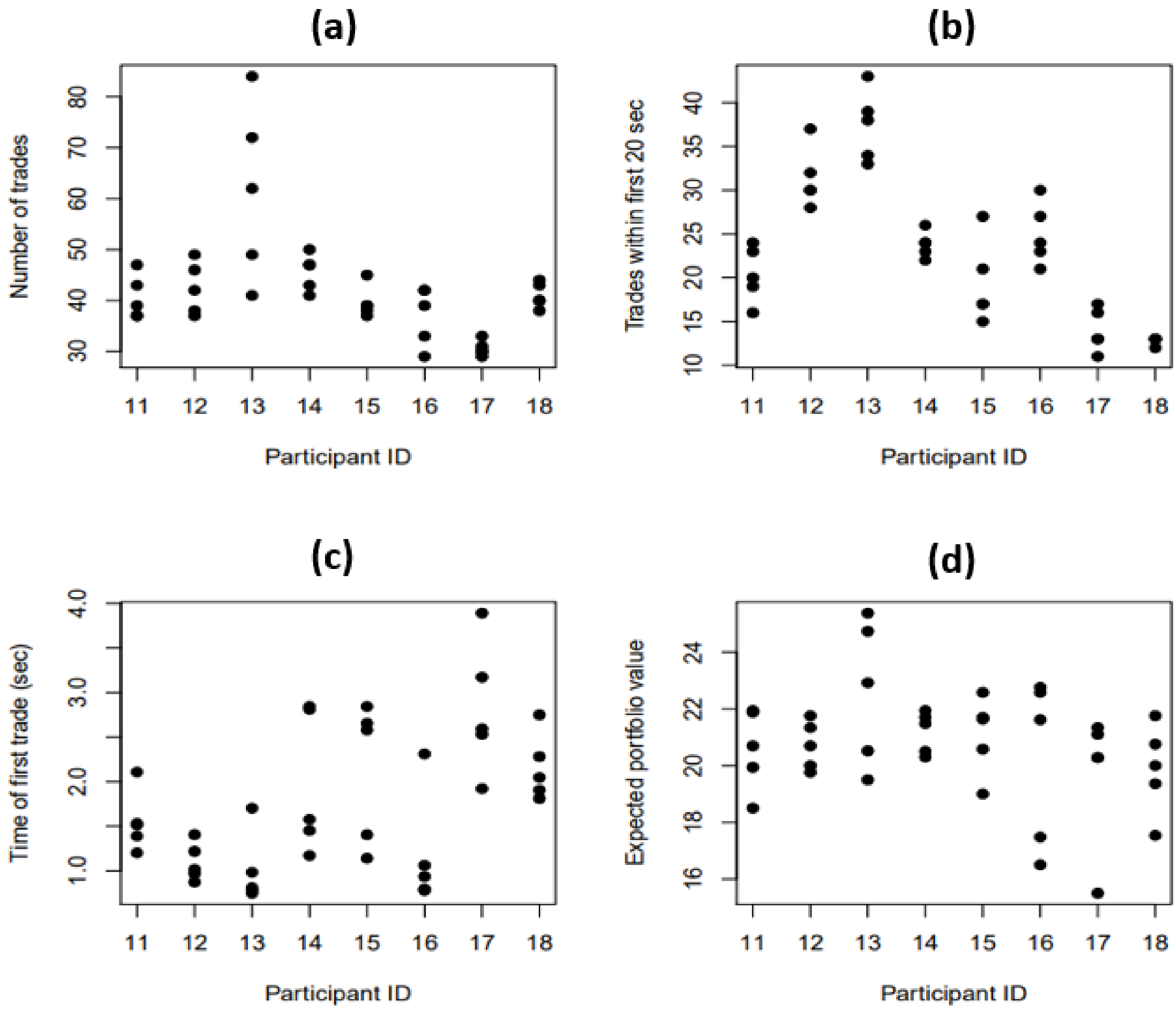


Figure 4.2 Graphs illustrating the individual variation in the four trading variables (y-axis) for eight participants (x-axis) from Session 1 for the five Stage 1 prediction markets. Panel (a) shows the number of trades per participant, (b) shows the number of trades executed in the first 20 seconds, (c) shows the time of the first trade and (d) shows the expected value of final portfolio holdings.

Figure 4.2 depicts the individual differences among traders. There are eight participants (coded by numbers ranging from 11 to 18), with five data points for each variable. The highest variance is seen in panel (b), where participant 13 trades substantially more than participant 18. On average, participant 13 trades 35 times whilst participant 18 trades 13 times. Panel (c) represents the time of the first trade. If a participant trades early, this time will be shorter; for participants that spread out their trades more evenly throughout the round, this time will be longer. Overall, Figure

4.2 shows that, while within-participant variance is relatively low, between-participant variation is relatively high.

Table 4.3 The effects of Session ID on the number of trades

Dependent Variable	Model(2)	Model(3)
	NumTrades20	FirstTrade
SessionID	-0.582**	0.189**
	(0.206)	(0.0675)
Constant	21.21***	1.514*
	(1.786)	(0.586)
Observations	480	480
R^2	0.755	0.640
Adj. R^2	0.693	0.550

Notes: Standard errors in parentheses. Significance at *.05, **.01, ***.001.

It is important to note that we find a statistically significant effect of Session ID on the total number of trades in the first twenty seconds and the time of the first trade at p<0.001 in models (2) and (3), respectively. We explored these effects further by looking at the coefficients from the statistical model. Table 4.3 displays the estimates for SessionID, used as a continuous variable, for models (2) and (3) from Table 4.2. The coefficients from both models suggest that when SessionID increases (from 1 to 12), participants' initial trades become slower in later sessions, i.e., the first transaction gets longer. This demonstrates that the timing of trades is affected by the experimental session and that each participant significantly affects trading. Because each subject participates only once in an experimental session, the models used for the ANOVAs are nested, and the model coefficients shown in Table 4.3 are difficult to interpret, without considering the individual-specific effects of participants in a session.

4.3 The Relationship Between Individual Trading Behavior and Participant Characteristics

We explore the experimental evidence on risk aversion and demographic characteristics of participants that could explain the variation in trading behavior. The variables used in our analysis are described in Table 4.4; Furthermore, Appendix E shows the demographic questionnaire and MPL risk lottery from Stage 3, and Appendix G shows two MPL risk aversion classifications.

Table 4.4 Description of variables for exploring what demographics drive trading

Variable name	**Description**
AvgNumTrades	The average number of trades per participant per market
AvgNumTrades20	The average number of trades in the first twenty seconds
AvgFirstTrade	The average time of the first trade executed in a market
AvgPortfolioValue	Average expected value of final portfolio holdings for each participant in a market
Gender	Categorical variable. Participant is either Female or Male
BusSchool	Categorical variable. Participant's field of study is from the Business School. We use Business School as our baseline as 57% of our cohort were recruited from the business school. Furthermore, we follow a similar approach adopted by Ackert et al. (1997) where the authors recruited only participants who have taken a statistics course
Age	Participant age in years
StatsCourse	Categorical variable. Whether a participant has taken a statistical course before
SafeChoices	The participant's risk attitude is defined by the number of A-choices, ranging from 0 to 10, where a higher number indicates more aversion to risk
Inconsistent	Inconsistent decisions are defined as switching more than once or making backward choices (switching in the opposite way)

We use OLS models to explore whether there is a relationship between Stage 3 demographic variables (gender, age, whether participants took a statistical course and if they are from the business school cohort as well as the number of safe and inconsistent choices each participant did in the lottery task) and the individual trading behavior variables we observed from Tables 4.2 and 4.3 (the averages of the number of trades per participant per market, number of trades within the first twenty seconds, time of the first trade executed in the market and the expected value of final portfolio holdings for each participant). The results are shown in Table 4.5. In Table 4.6, we use further OLS models to explore relations between trading characteristics and the final portfolio value.

Table 4.5 Characteristics that drive trading and the relationship between risk aversion and trading behavior

Dependent Variable	(1)OLS AvgNumTrades	(2)OLS AvgNumTrades20	(3)OLS AvgFirstTrade	(4)OLS AvgPortfolioValue
Gender	-0.267	0.589	-0.104	0.0805
	(1.190)	(1.204)	(0.309)	(0.281)
Age	-0.191	-0.191	0.0115	-0.0353
	(0.172)	(0.174)	(0.0447)	(0.0407)
StatsCourse	-1.443	-3.464*	0.303	-0.123
	(1.523)	(1.540)	(0.395)	(0.360)
BusSchool	0.892	0.436	0.167	-0.183
	(1.240)	(1.255)	(0.322)	(0.293)
SafeChoices	-0.0237	-0.0872	0.0327	-0.00409
	(0.327)	(0.331)	(0.0849)	(0.0773)
Inconsistent	-2.543	-0.582	-0.0574	-0.537
	(1.549)	(1.567)	(0.402)	(0.366)
Constant	46.53***	26.80***	1.457	21.65***
	(4.617)	(4.671)	(1.197)	(1.091)
Observations	96	96	96	96
R^2	0.064	0.073	0.016	0.043
Adj. R^2	0.001	0.010	-0.050	-0.021

Notes: Standard errors in parentheses. Significance at *.05, **.01, ***.001.

Table 4.6 The effect of different trading metrics on participants' portfolio value

A set of OLS models are implemented with average portfolio value as our dependent variable for 96 participants.

Dependent Variable	(1)	(2)	(3)	(4)
AvgNumTrades	0.0342		0.0209	
	(0.0238)		(0.0275)	
AvgNumTrades20		0.0379	0.0238	
		(0.0234)	(0.0292)	
AvgFirstTrade			-0.0911	-0.0398
			(0.0948)	(0.104)
Constant	19.18***	19.84***	20.78***	19.35***
	(0.985)	(0.479)	(0.245)	(1.060)
Observations	96	96	96	96
R^2	0.021	0.027	0.010	0.034
Adj. R^2	0.011	0.017	-0.001	0.003

Notes: Standard errors in parentheses. Significance at *.05, **.01, ***.001.

Result 2: Individual differences in trading behavior are not linked to any individual characteristics.

Whilst ANOVA results from Table 4.2 show that there are individual differences in trading behavior, results from Tables 4.5 show that those differences are not linked to Stage 3 variables. Looking at whether demographic characteristics are correlated with trading behavior, we find that none of them, including gender and whether participants had previously taken a statistical course, show a statistically significant effect. We find no evidence that risk aversion is related to trading behavior, and also find no effect of trading characteristics on the final portfolio values from Table 4.6.

4.4 The Impact of Trading Behavior on Market Efficiency

In this section we aim to answer our third hypothesis by investigating whether market accuracy depends on the number and timing of trades in a market. We explore this relationship using absolute error and the absolute error of logit transformed probabilities as dependent variables to quantify market accuracy, and independent variables similar to the ones used in Section 4.2 and 4.3 to quantify behavior in a market. The behavioral variables are shown in Table 4.7, the results are shown in Table 4.8 for absolute error and Table 4.9 for the absolute error of logit transformed probabilities.

Table 4.7 Description of variables for exploring how trading behavior affects market efficiency.

Variable name	**Description**
TotalTrades	The total number of trades for both participants per market.
TotalTrades20	The total number of trades for both participants per market in the first twenty seconds.
TotalFirstTrade	The time of the first trade executed by a participant in a market.
TotalPortfolioValue	The total expected value of final portfolio holdings for both participants per market.

Trading performance and market efficiency

Table 4.8 The dependence of market efficiency on trading behavior

A set of OLS models are implemented with absolute error as our dependent variable on 240 markets

Dependent Variable	(1)	(2)	(3)	(4)	(5)
TotalTrades	0.00125			0.00142	
	(0.000891)			(0.000927)	
TotalTrades20		-0.000290		-0.000398	
		(0.000907)		(0.000993)	
TotalFirstTrade			0.0105	0.0103	
			(0.0109)	(0.0116)	
TotalPortfolioValue					-0.0358***
					(0.00290)
Constant	0.0464	0.160***	0.133***	0.0328	1.624***
	(0.0735)	(0.0368)	(0.0183)	(0.0807)	(0.120)
Observations	240	240	240	240	240
R^2	0.008	0.000	0.004	0.014	0.391
Adj. R^2	0.004	-0.004	-0.000	0.001	0.388

Notes: Standard errors in parentheses. Significance at *.05, **.01, ***.001.

Table 4.9 Robustness of results using an alternative error measure

A set of OLS models are implemented with absolute error of logit transformed probabilities as our dependent variable on 240 markets.

Dependent Variable	(1)	(2)	(3)	(4)	(5)
TotalTrades	0.00743 (0.00519)			0.00917 (0.00539)	
TotalTrades20		-0.00416 (0.00528)		-0.00518 (0.00578)	
TotalFirstTrade			0.0663 (0.0637)	0.0544 (0.0674)	
TotalPortfolioValue					-0.133*** (0.0199)
Constant	0.552 (0.428)	1.324*** (0.214)	1.062*** (0.106)	0.534 (0.470)	6.625*** (0.820)
Observations	240	240	240	240	240
R^2	0.009	0.003	0.005	0.017	0.158
Adj. R^2	0.004	-0.002	0.000	0.005	0.154

Notes: Standard errors in parentheses. Significance at *.05, **.01, ***.001.

Result 3: Individual trading variables do not influence market accuracy.

Results from Models (1) to (4), from table 4.8, indicate that there are no significant effects from any of the behavioral variables on market efficiency. In model (5) from Table 4.8, we analyze if there is a relationship between the final portfolio holdings of participants and the absolute error. We find that final portfolio values are closely correlated with market accuracy. The direction of the effect signifies that a lower error is correlated with higher portfolio values. This is expected because the market scoring rules relate purchased assets with the pricing, and thus the accuracy. When participants' trades were well aligned with their signals, the markets were more accurate. Our results also remain robust when using the alternative error measure in Table 4.9.

Trading performance by signal constellations

The findings from Tables 4.8 and 4.9 prompted an in-depth examination of the various trading parameters to understand whether the insignificant effects observed depend on the constellation of signals received by the participants. We perform further OLS analysis on the number of trades, number of trades in the first twenty seconds and timing of the first trade for signal constellations where participants receive the same signals (BB and RR), and where they receive different signals (RB). The results are shown in Table 4.10.

Table 4.10 The dependence of the number of trades on signals

	BB & RR		RB	
Dependent Variable	(1) AbsoluteError	(2) AbsoluteLogitError	(3) AbsoluteError	(4) AbsoluteLogitError
TotalTrades	0.00228* (0.000885)	0.0126* (0.00523)	-0.00856** (0.00298)	-0.0436* (0.0177)
Constant	-0.0498 (0.0731)	0.177 (0.432)	0.896*** (0.244)	4.537** (1.454)
Observations	192	192	48	48
R^2	0.034	0.029	0.152	0.116
Adj. R^2	0.029	0.024	0.134	0.097

Notes: Standard errors in parentheses. Significance at *.05, **.01, ***.001.

Results show that the effect of the total number of trades on market accuracy depends on the signal constellations. We tested each trading variable independently per signal constellation (results not shown) but found no effect except across the total number of trades. When signals are the same for both trades (RR and BB), more trades mean more efficiency (larger error), when the signals are different (RB) more trades mean less efficiency (smaller error). This is also illustrated in the scatter plots in Figure 4.3(a) and 4.3(b), respectively.

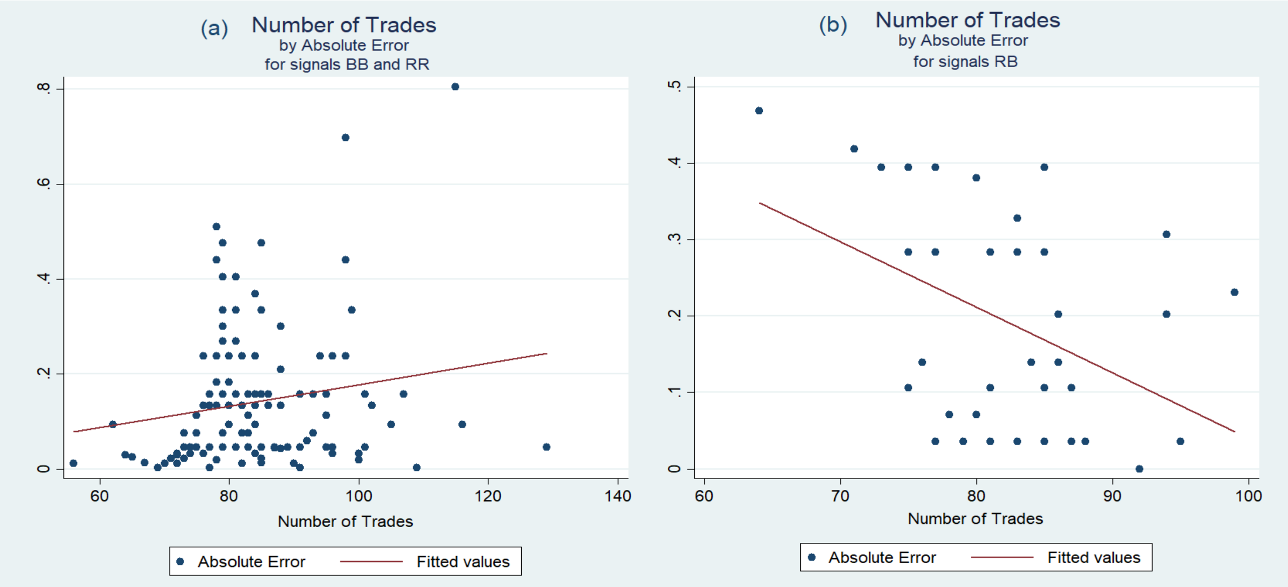


Figure 4.3 Scatter plots showing the absolute errors (y-axis) by total number of trades (x-axis) depending on different signal constellations provided to the participants. Signals provided are BB= Blue, Blue , RB= Red, Blue and RR = Red, Red.

4.5 Final Holdings in the Stage 2 Markets

In this section, we follow up on the findings from Chapter 3.8 about errors in the DMD setting in our Stage 2 markets. When both signals are blue, we found that higher errors might be driven by overpricing. As such, we further test and investigate how participants behavior differ between the Stage 2 settings. We confine this analysis on the final holdings of participants in the markets.

Table 4.11 Description of variables for understanding traders trading preferences.

Variable name	Description
RedHoldings	The total final holdings of Red Type assets of all participants in the two markets (for Urn 1 and Urn 2) a group can trade in, for a total of 192 groups.
BlueHoldings	The total final holdings of Blue Type assets of all participants in the two markets (for Urn 1 and Urn 2) a group can trade in, for a total of 192 groups.
AbsoluteDiff	The absolute difference between the total assets from market 1 and the total assets from market 2 for the 192 groups.

Table 4.12 Do participants trade certain assets in a specific setting?

Three sets of OLS models are implemented with RedHoldings, BlueHoldings and AbsoluteDiff as our dependent variables on 192 groups. We include a dummy variable for DMD and DMS with the control treatment PM2 as the base category. From Model (2), we test the difference between DMD and DMS and we find a statistically significant at 5.047, p<0.01. The difference between DMD and DMS in Model (3) is statistically significant at -63.70, p<0.001.

Dependent Variable	(1) RedHoldings	(2) BlueHoldings	(3) AbsoluteDiff
DMD	2.422	-5.484**	67.88***
	(1.808)	(1.820)	(5.591)
DMS	2.797	-0.438	4.172
	(1.808)	(1.820)	(5.591)
Constant	73.06***	74.53***	20.66***
	(1.279)	(1.287)	(3.954)
Observations	192	192	192
R^2	0.015	0.056	0.495
Adj.R^2	0.004	0.046	0.490

Notes: Standard errors in parentheses. Significance at *.05, **.01,***.001. PM2 is used as a baseline across all our models.

In the markets, the total final holdings were on average 75 (Std. Dev.=10.25) for Red Type assets, and 73 (Std. Dev.= 10.54) for Blue Type assets. From Model (2), participants avoid Blue Type assets in DMD but not in other settings (there is no such effect from Model (1) for Red Type assets). We do not see this effect from the results of our Model (1). Under the DMD setting, one market is frequently avoided whilst the other is favored. From the results of our models, we can clearly see participants behavior in the DMD settings differ from other settings. A small absolute difference in our Model (3) for DMS indicates that participants in a group spread their investments (final holdings) evenly across the two markets. Meanwhile in our DMD setting, the large absolute difference indicates that the final holdings are largely in one of the two markets.

4.6 Conclusion

In this chapter we investigated if there are any behavioral attributes associated with trading performance in our Stage 1 prediction market experiment. Our main finding (Result 1) is that there is evidence for differences in individual trading behavior. A large portion of variation in trading behavior is attributed to between-subject variation indicating that trading significantly varies according to each participant. The variable with largest between-subject variation is the number of trades within the first 20 seconds of a round. The smallest between-variation is observed for expected final portfolio value. This indicates that trading behavior such as trading activity does not have a strong effect on portfolio values, which is further supported by the analyses presented in Section 4.3 (Table 4.6) on the relation between trading behavior and final portfolio value.

Our second result (Result 2) is that despite individual differences in trading behavior being detected, there is no evidence for any relation to Stage 3 behavioral attributes. Our final and third result finds no evidence for any effect of trading behavior on market efficiency. However, after an in-depth analysis stemmed from the third result, we find that the total number of trades on market accuracy is dependent on the signal constellations. When the signal constellations are the same, more trades mean more market efficiency and when the signals are different, more trades mean less efficiency.

Furthermore, in Section 4.5, we observe that participants' behavior in the DMD settings differs from the other settings. Participants avoid holding contracts for Blue Type urns and they place their final holdings largely in one market rather than spreading it across two markets.

Overall, our results from the Stage 1 markets provide evidence for individual differences but reveal little about the reasons behind these differences. The individual variation we observed could in principle have an effect in markets. While we do not observe any effect of trading behavior on market efficiency, such effects could have an effect in markets with a design that differs from the design we used in our experiments. Future studies could use a more detailed assessment of behavioral attributes on trading behavior to identify reasons behind individual differences in trading behavior and investigate circumstance in which behavioral variation has a more substantial impact on the performance of individual traders and the functioning of markets. Contrasting the results from Stage 1, our analysis of the Stage 2 markets shows that the decision rule does have an impact on the traders' behavior in the market, which in turn has an effect on pricing and error.

Chapter 5: Discussion and Conclusion

Decision markets have been developed in theory, but there have been very few applications. Significant empirical research is required to investigate the fundamentals and functioning of such markets and bring them into practice. As such, our main research question was to examine whether decision markets aggregate information and if their predictive performance matches the performance of a prediction market. Five hypotheses were stemmed from the main research question to answer and explore the performance and characteristics of such markets.

This thesis experimentally investigated the impact of information aggregation on decision-making. The aim was to provide and implement an experimental demonstration of decision markets through an analogue of Plott and Sunders's (1988) prediction market experiment. Trader characteristics, behaviors and incentives that could have been associated with information aggregation were also examined to provide us with a better understanding of how such characteristics affected information aggregation.

In Chapter 2, I describe the design of the laboratory experiment used in this thesis to study how different decision rules in decision markets affect information aggregation. Through the design of the market experiment, we investigate whether the predictive performance of prediction markets can be matched by a decision market, and if so, whether it is sensitive to the decision rules used. We simulate a situation where a decision-maker must make a choice over possible decisions to execute, where the possible decisions will impact the probability of achieving a desired outcome. Participants enter the market as pre-established experts to trade their expectations about the outcome of future events under three different settings, each of which employs a different decision rule. Participants are then rewarded for their accurate predictions based on the decision scoring rule used in the setting.

Chapters 3 and 4 are dedicated to describing and analyzing the experimental results from our decision market experiment. In Chapter 3, to answer our main research question, we address two hypotheses: whether final market prices are affected by participants' private signals and if market errors are affected by the decision rule applied. We do so by implementing a set of OLS models using two different error measures.

From our first hypothesis, we find that market prices are dependent on the private signals given to participants. This signifies that the signal constellations are the primary determinant of final market prices. According to the findings from our second hypothesis, we can confirm that market errors depend on the decision rule applied. Errors are largest in our deterministic setting, demonstrating that the decision rule's incentive compatibility is important for information aggregation. Furthermore, we find that in the deterministic setting, prices are higher. This indicates that the high errors might be due to the inflated prices in the settings (particularly when the signal constellations are BB or RB).

Theoretical and empirical research in financial markets support the concept of "Heterogeneity in Belief," which can contribute to determining final market prices. Because market participants come from diverse backgrounds (education, life experience, ethnicity, etc.), they may interpret the same signal differently, resulting in different valuations for the same assets. For example, individuals from an Asian descent believe that red is symbolizes luck and happiness. This might be one of the reasons why most participants tend to hold on to Red Type assets (given that 54% of our sample were from an Asian ethnicity). To some degree, the experiments “manipulate” the dispersion of belief of market participants by providing them with signals. Nonetheless, we do not know how participants themselves are interpreting these signals, and how they learn from the market. Even participants who receive a red signal could be heterogeneous in a sense that different participants might interpret the different posteriors based on their different life experiences. Such heterogeneity in behaviour poses a general problem for experiments on believe elicitation and could be explored more in the future.

Chapter 4 builds up on the results of Chapter 3. We use data from the demographic questionnaire and the lottery task to examine participants' risk preferences and explore whether there are behavioral attributes linked to their trading performance from our Stage 1 prediction market experiment. We do this by addressing three hypotheses.

In our first hypothesis, we explore whether there are any individual differences in participants trading behavior. We find that trading varies greatly depending on each participant. In our second hypothesis, we follow up by investigating whether the observed individual differences in trading behavior are linked to the individual characteristics from the questionnaire and lottery task. We find no evidence for any relation to the Stage 3 behavioral attributes. A factor potentially explaining this lack of evidence is that our cohort may be relatively homogeneous, with participants being mostly university students in a similar age range.

The final hypothesis extends the second hypothesis' findings by investigating whether individual variations influence market accuracy. We find no evidence that trading behavior has any influence on market efficiency. However, after more investigation, we discover that the total number of trades on market accuracy is reliant on the signal constellations. When the signal constellations are the same, indicate higher market efficiency; when the signal constellations are different, more trades mean lower market efficiency. We also notice that in the deterministic environment, participants avoid retaining contracts for Blue Type urns, and as a result, they concentrate their final holdings in one market rather than distributing them over two markets.

In conclusion, theory has shown that decision markets are strategically and mechanistically complex. This thesis investigated how humans use information in individual decision making and strategic interactions to reach a forecast about future events. Overall, our results show that decision markets work in aggregating private information and that the incentive compatibility of the decision rule matters for information aggregation. Participants had an incentive to truthfully reveal their beliefs. Their beliefs were either inferred from their decisions (upon receiving private signals) or from their own assessment. Despite the successful performance of decision markets, they still do not outperform those of prediction markets. Upon exploring behavioral attributes that might be linked to individual trading performance, we discovered that the decision rules do have an impact on participants' behavior in the market, which in turn manifested itself into the assets pricing and errors.

In the research I conducted, participants beliefs were elicited from their decisions about predicting the outcome of events represented by one and two urns. A first direction in which my research could be extended is by introducing three states of the world rather than two. Since our decision market setup produced sufficient performance in simple scenarios, this simple proof-of-principle could be extended to an analogue of Plott and Sunders's prediction market experiment to a conditional forecasting scenario with three mutually exclusive outcomes. This could be done by adjusting the experimental design to test more complex scenarios that introduce three urns instead of two urns.

A second interesting aspect would be to alter the ball compositions to observe more actions. Instead of having an urn that has ten balls with one private signal, one could have the urn consisting of twenty balls with two private signals (or even ten balls with two signals). This would allow us to extract more analysis from the different actions that traders will execute and test whether information will be aggregated as those from the simple scenarios. A third direction is to construct the experiment so that one person receives more private information than the others. Under such a setting, we would be able to observe whether decision markets perform similarly if 'one' expert holds more private information than those of other traders and whether the market would converge to the real state of the world. Moreover, in such a setting, decision markets are likely to outperform voting where each individual has the same impact on the final decision.

A fourth and final intriguing feature would be a design similar to that of Page and Siemroth (2017), in which participants incur a cost to receive more private signals. This would allow us to see how much individuals value information. However, given the complexity of decision markets, it is critical to ensure that, instructions are well described and that the experimental design is as simple as possible.[19]

We believe that results from our research offers the potential to improve any evidence-based and forecast-based decision-making where desirable. It could also be used as an implementation or comparison to other decision-making mechanisms, such as voting. Because decision markets can be seen as voting mechanisms that connect a vote for an action with a forecast of its consequences, this relationship between voting

[19] Despite this, our results revealed that under more complex settings (stochastic), we get a better performance.

and forecasting in decision markets might offer exciting opportunities for customizing the design of voting mechanisms. An example would be the well-known Condorcet Jury Theorem, where the assumption is that voters vote informatively based on their signal and that individuals behave in the collective decision precisely as they would if they were voting alone.

Appendices

Appendix A: Control Questions

Prediction Market (PM1)

1) At the beginning of the round, **before** anyone has received their signals, what's the probability that the bucket will be a Blue Type?

a) 50%

b) 100%

c) 80%

2) Suppose your private signal is a Red ball, what can this tell you about the bucket?

a) It is a Red Type

b) It is a Blue Type

c) It can be either a Red or Blue Type, but with a higher probability of being a Red Type

d) It can be either a Red or Blue Type, but with a higher probability of being a Blue Type

3) If you have bought 6 Blue Type assets and 12 Red Type assets, and the bucket outcome turned out to be a Blue Type, what will be your total payoffs?

a) 6 e-dollars

b) 12 e-dollars

c) 18 e-dollars

d) 0 e-dollars

4) One ball was drawn as your private signal and then put back into the bucket. How many balls are now in the bucket?

a) 10

b) 9

c) 11

d) 4

Two Simultaneous Prediction Markets (PM2)

1) What is the main difference between this part of the experiment and the previous part (part 1)?

a) There will be one bucket

b) There will be two buckets

c) No difference

2) Are the two buckets always the same color?

a) Always the same

b) Always different

c) Would not know

3) Suppose in round 3, you have earned 18 e-dollars. What is the total amount of e-dollars that you can use to trade in next round?

a) 18

b) 20

c) 38

4) Suppose the current market price of a Blue Type asset is 0.6. In order to increase that price, what would you do?

a) Buy more Blue assets

b) Buy more Red assets

c) Buy both Blue and Red assets

d) Do nothing

Deterministic Decision Market (DMD)

1) What is the main difference between this part of the experiment and the previous part (part 1)?
   a) There will be one bucket.
   b) There will be two buckets.
   c) No difference.

2) Are the two buckets always the same color?
   a) Always the same
   b) Always different
   c) Would not know

3) If the Red assets in Bucket 2 had the highest final market price, what's the probability of the computer selecting it to generate your payoff?
   a) 50%
   b) 100%
   c) 25%
   d) Unknown

4) If the computer has selected bucket 2 to be revealed, will the computer also reveal bucket 1?

   a) Yes

   b) No

   c) Maybe

Stochastic Decision Market (DMS)

1) What is the main difference between this part of the experiment and the previous part (part 1)?
   a) There will be one bucket
   b) There will be two buckets
   c) No difference

2) Are the two buckets always the same color?
   a) Always the same
   b) Always different
   c) Would not know

3) If the Red assets in Bucket 2 had the highest final market price, what's the probability of the computer selecting it to generate your payoff?

   a) 50%

   b) 100%

   c) 75%

   d) Unknown

Appendix B: Stage 1 Instructions

Welcome to the Lab

Welcome and thank you for registering in today's experimental session!

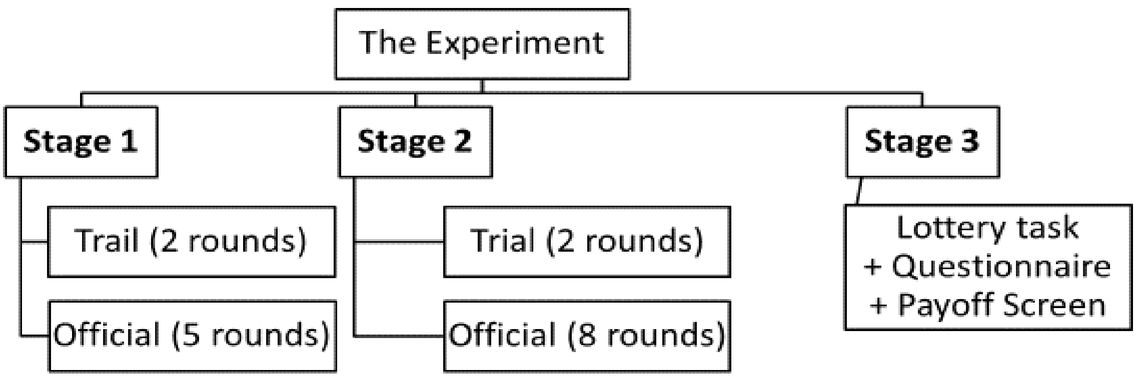
- You can choose to stay or leave the experiment by the end of the instructions.
- Please follow the instructions carefully to smoothly run the experiment.
- Please use the computer only for participating in the Experiment. **DO NOT OPEN ANYTHING ELSE.** Do not open or close any programs, unless instructed.
- From now on, please do not communicate with other participants.

Note: Please raise your hand if you have any questions now or later. An Experimenter will come and assist you.

Remember: *Similar rules of behaviour as during examinations apply to today's experimental session.*

Outline of today's experiment

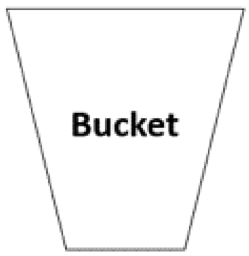
You will participate in **three different Stages** during today's experimental session:



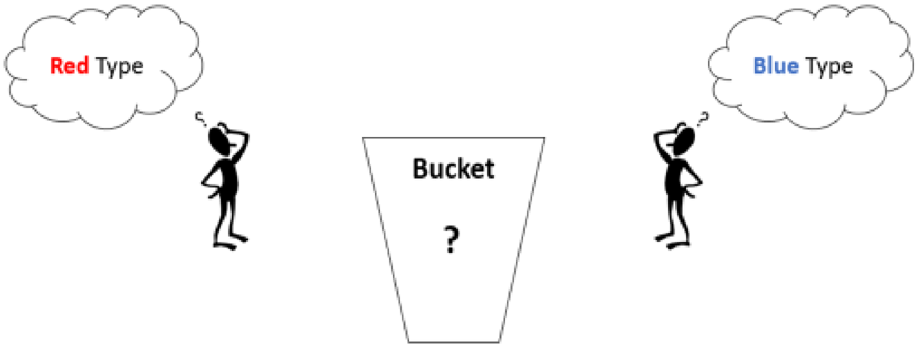
The Experiment

- The aim of this experiment is to study how you use information about **uncertain events to make decisions.**
- The uncertain events that you will be faced with are represented by **BUCKETS.**

In each round, there will be a bucket:

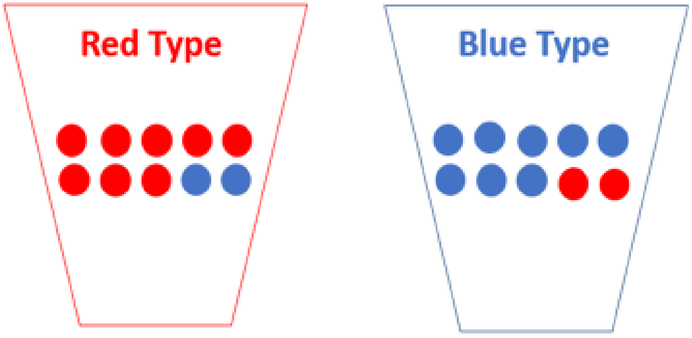


This bucket has a **type.** The type of this bucket is unknown. It can be either of a **Red Type** or a **Blue Type**.

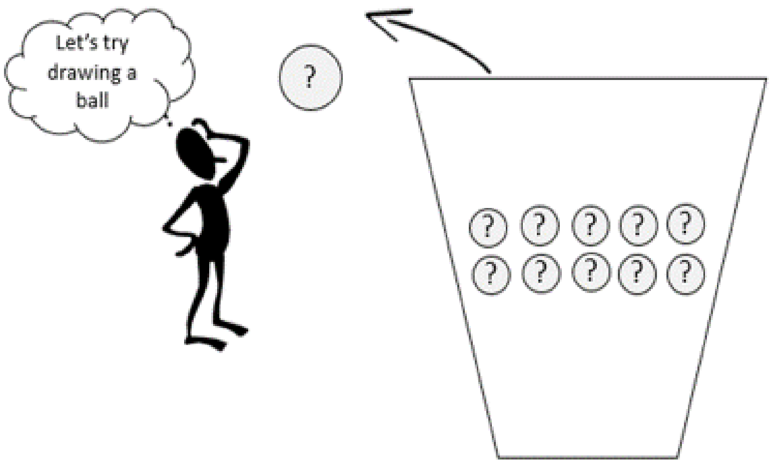


Both types have <u>**10 balls each**</u>

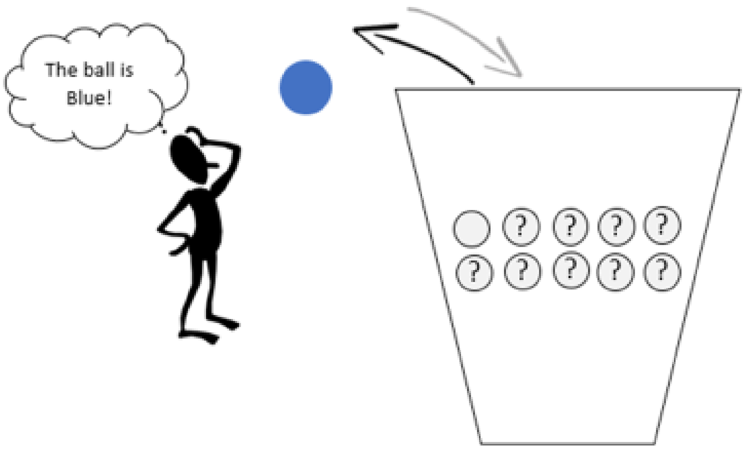
- A Red Type bucket contains: 8 red balls and 2 blue balls.
- A Blue Type bucket contains: 8 blue balls and 2 red balls.



- You will get the chance to have **one** ball drawn **randomly** from the bucket.
- You will see the color of this ball.
- The color of this ball is your **private information**.
- This private information will give you an idea about the type of the bucket.
- Other participants will not know your private information and you will not know theirs.



Example: The outcome of your private draw from the bucket was **blue.** After you have seen the color of your drawn ball, **it will be put back into the same bucket.**



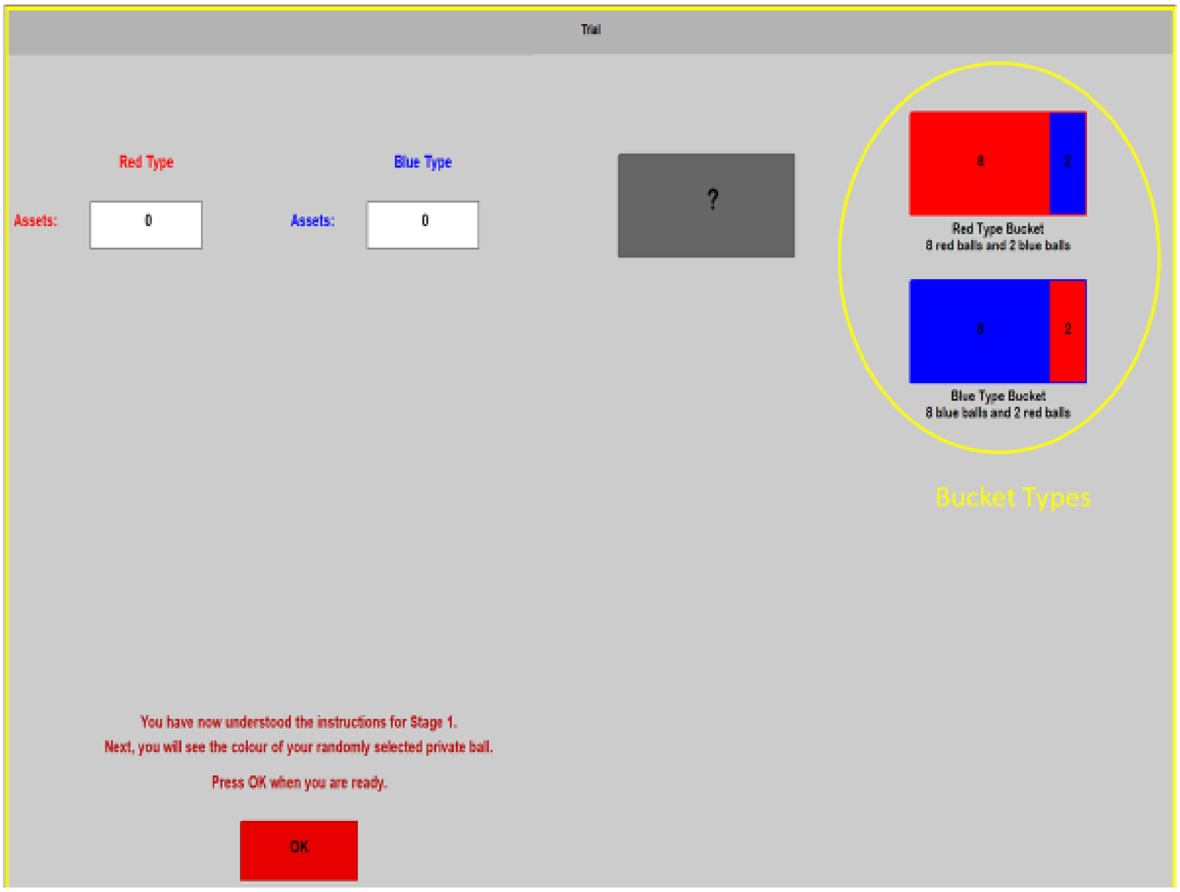
Stage 1

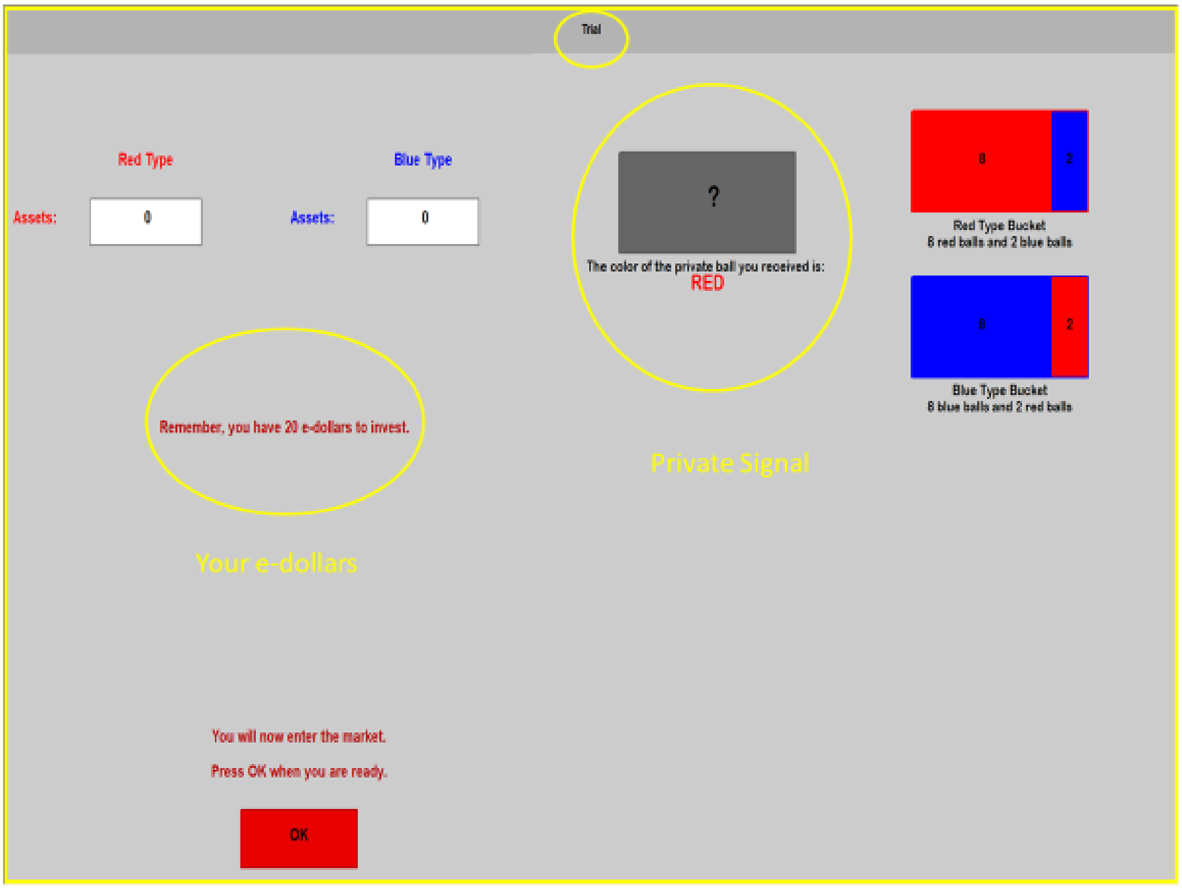
After seeing the color of your drawn ball, you will enter a **market.** This market will allow you to **trade assets** on the type of **one** bucket.

- You will have 20 e-dollars to trade/purchase assets with.
- The 20 e-dollars will expire at the end of **each** round.
- Each round will last 1 minute. You have 5 rounds.
- Practice rounds are **not paid** but you will get to see your payoffs at the end to give you an idea of how your payoffs will look like.

In each round:

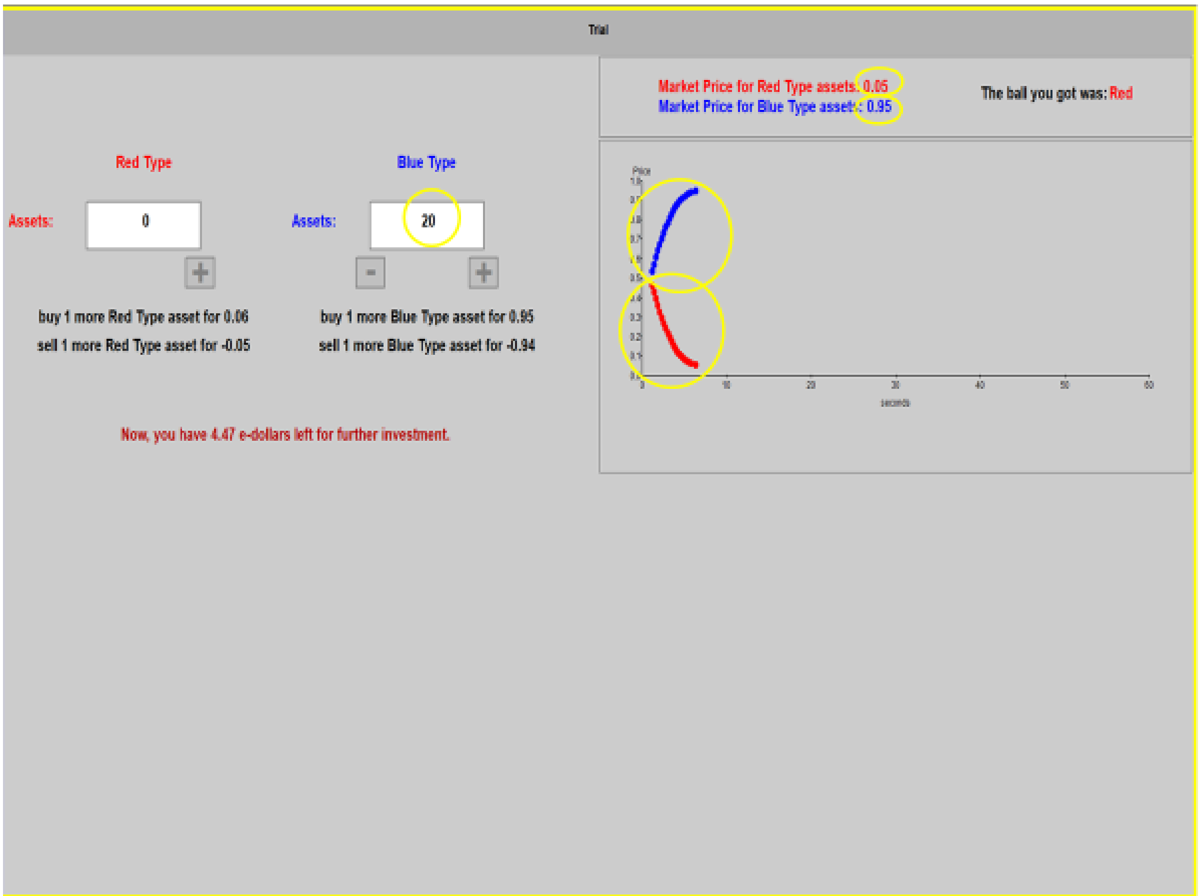
1. There will be **ONE** bucket.
2. The type of this bucket is unknown. It could be of a Blue Type or a Red Type.
3. Whether it's a "Blue Type" or a "Red Type" is selected randomly by the computer.
4. The chance that the "Blue Type" gets selected is **50%**. The chance that the "Red Type" gets selected is also **50%**.
5. You will be randomly grouped with **one** other person. You will not be directly trading with this person but the money you earn will depend on what you decide and what the other person in your group decides as well.


Trial
Red Type
Blue Type
Assets:
0
Assets:
0
?
8
2
Red Type Bucket
8 red balls and 2 blue balls
8
2
Blue Type Bucket
8 blue balls and 2 red balls
Bucket Types
You have now understood the instructions for Stage 1.
Next, you will see the colour of your randomly selected private ball.
Press OK when you are ready.
OK


Trial
Red Type
Blue Type
Assets:
0
Assets:
0
?
The color of the private ball you received is:
RED
8
2
Red Type Bucket
8 red balls and 2 blue balls
8
2
Blue Type Bucket
8 blue balls and 2 red balls
Remember, you have 20 e-dollars to invest.
Private Signal
Your e-dollars
You will now enter the market.
Press OK when you are ready.
OK


Trial
Market Price for Red Type assets: 0.50
Market Price for Blue Type assets: 0.50
The ball you got was: Red
Market price
Private Signal
Red Type
Blue Type
Assets:
0
Assets:
0
buy 1 more Red Type asset for 0.52
sell 1 more Red Type asset for -0.48
buy 1 more Blue Type asset for 0.52
sell 1 more Blue Type asset for -0.48
Asset price
Now, you have 20.00 e-dollars left for further investment.


Trial
Market Price for Red Type assets: 0.86
Market Price for Blue Type assets: 0.14
The ball you got was: Red
Red Type
Blue Type
Assets:
13
Assets:
0
buy 1 more Red Type asset for 0.87
sell 1 more Red Type asset for -0.86
buy 1 more Blue Type asset for 0.14
sell 1 more Blue Type asset for -0.13
Now, you have 10.86 e-dollars left for further investment.


Trial
Market Price for Red Type assets: 0.05
Market Price for Blue Type assets: 0.95
The ball you got was: Red
Red Type
Blue Type
Assets:
0
Assets:
20
buy 1 more Red Type asset for 0.06
sell 1 more Red Type asset for -0.05
buy 1 more Blue Type asset for 0.95
sell 1 more Blue Type asset for -0.94
Now, you have 4.47 e-dollars left for further investment.
seconds


Trial
Market Price for Red Type assets: 0.15
Market Price for Blue Type assets: 0.85
The ball you got was: Red
Red Type
Blue Type
Assets:
11
Assets:
23
buy 1 more Red Type asset for 0.16
sell 1 more Red Type asset for -0.14
buy 1 more Blue Type asset for 0.86
sell 1 more Blue Type asset for -0.84
Now, you have 0.71 e-dollars left for further investment.
e-dollars left
seconds
Time

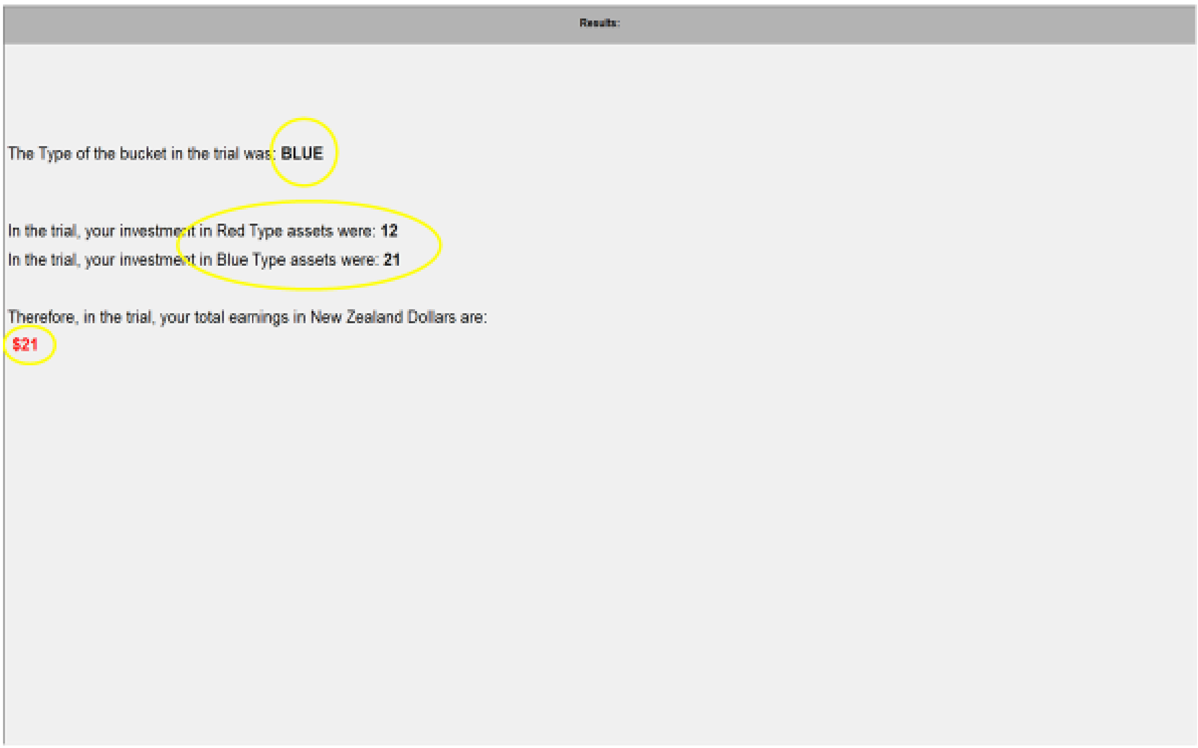
Stage 1-Payoff

After all trading rounds, the computer will **randomly** select **one round** and reveal the type of the bucket for that round.

- If the true type of the bucket turned out to be Red, you will be paid **1 e-dollar for every Red asset you purchased; 0 e-dollars for Blue.**
- If the true type of the bucket turned out to be Blue, you will be paid **1 e-dollar for every Blue asset you purchased; 0 e-dollars for Red.**

Your rewards for Stage 1, based on the decisions you made for the selected round, will be revealed to you at the end of today's experiment.

Payoff Example



Ethics

Ethics considerations

Participation is entirely voluntary and anonymous.

Earlier on:

When we invited you, we sent you a link to the **PARTICIPANT INFORMATION SHEET** with details of the experiment. You also have a copy of this on your desk.

Today:

Now you can choose to TAKE PART OR LEAVE the Experiment

If you agree to participate, please sign one copy of the **PARTICIPANT CONSENT FORM** for the experimenter to collect at the end of the session (one copy is for you to keep).

If not, please raise your hand so that the experimenter can assist you with collecting your show-up fee of **$10** and with leaving the experimental lab.

Remuneration

Besides the **$10** show up fee, you will earn:

- Minimum of **$25** if you decide to stay until the end of the experiment and,
- Up to **$50** depending on your performance in Stage 1, Stage 2 and Stage 3.

This means you can earn a maximum of $50 in today's experiment. You will receive your payment at the end of the experiment when your computer terminal number is called out.

Please bring the signed consent form to the experimenter when your computer terminal number is called out.

Before we get started:
Some 'House-Keeping' Rules

Please use computers only as instructed.

- Do not start or end any programs, unless told to do so.
- Do not change any settings.

Please only use the material provided.

***<u>Please note</u>**: If you have any questions either now, or during the experiment, raise your hand, and an experimenter will come and assist you privately.*

Instructions for Stage 1 is now over.

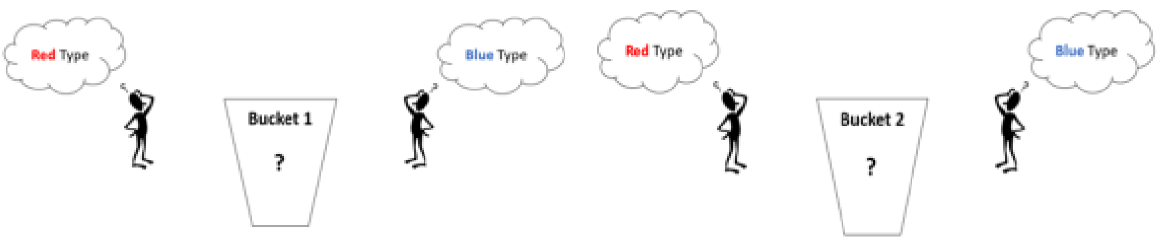
We will start off with a few control questions to help test your understanding of the instructions.

Appendix C: Stage 2 Instructions

We will now go through the instructions for Stage 2.

Stage 2

- You will participate in similar decision tasks.
- Instead of predicting the outcome of 1 bucket, you will be asked to predict the outcome of **2 buckets.**



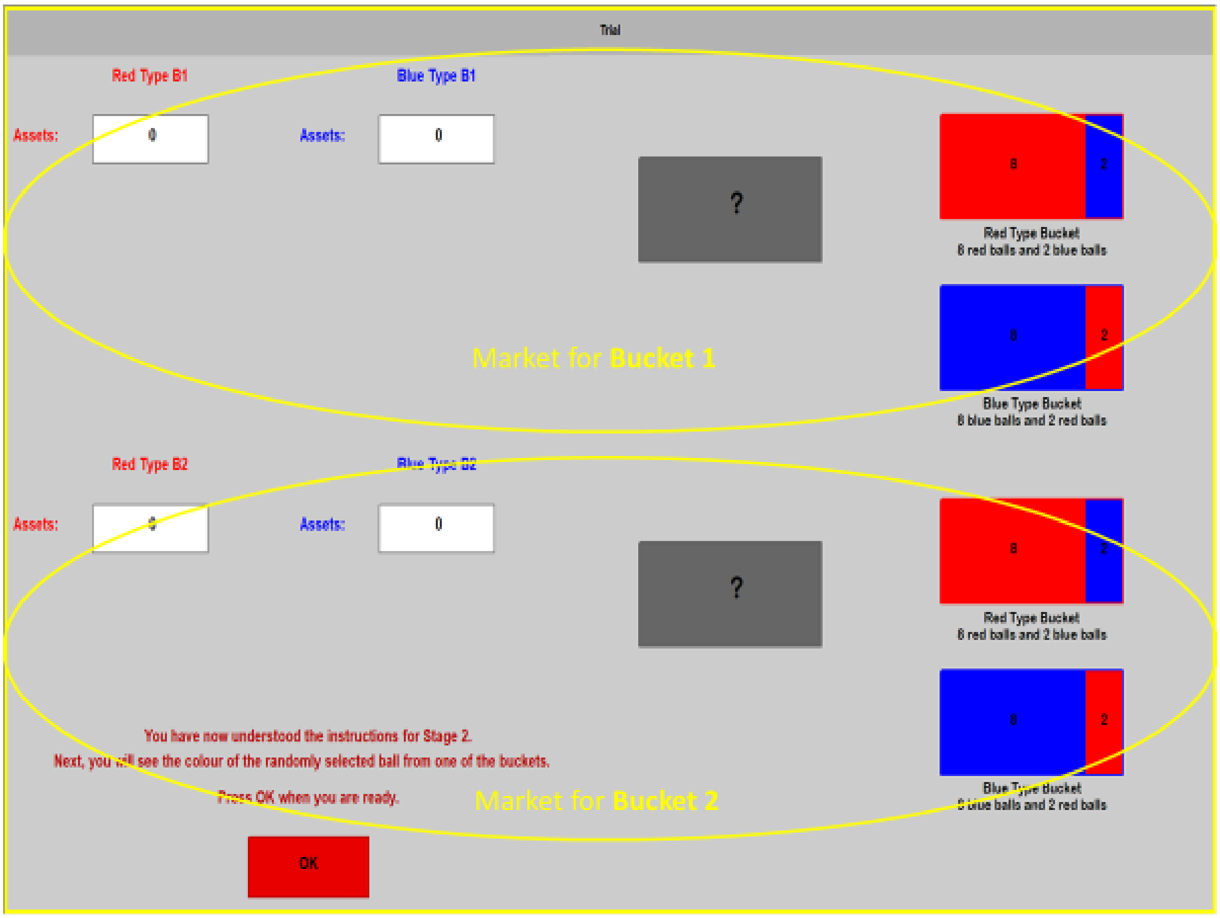
Stage 2

- You will get the chance to have one ball drawn randomly from **ONE** of the buckets.
- The computer will select which bucket you get your signal from + what ball color you get **randomly**.
- Despite your private signal is *only* from ***one*** of the buckets, you will get the chance to trade on *both buckets* in the market.
- The type of both buckets are unknown. They will be selected **randomly** by the computer. The chance that either type gets selected is 50%.
- You will randomly be allocated with **3 other people** with you in the group.

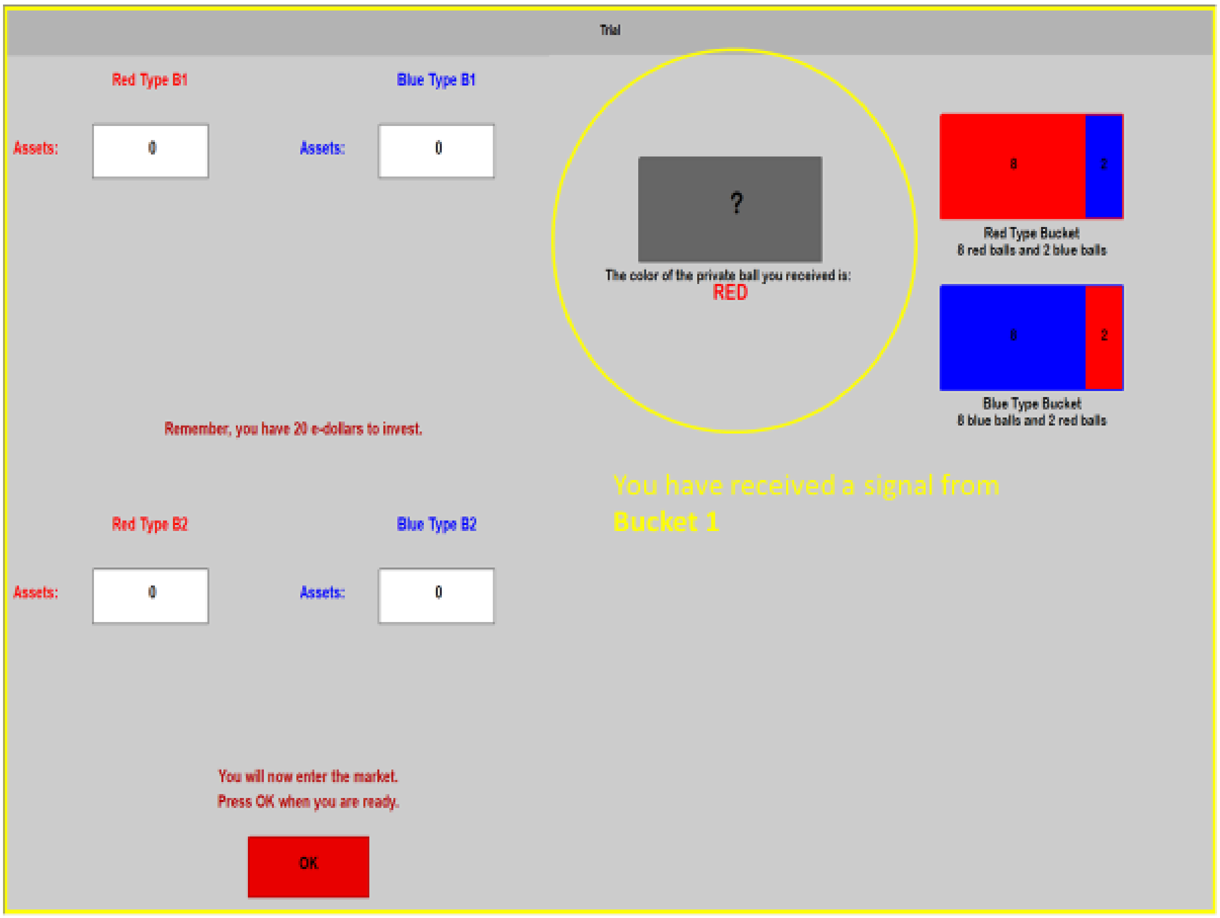
Stage 2

After seeing the color of your drawn ball, you will enter **two markets.** These two markets will allow you to trade assets on the type of **two buckets.**

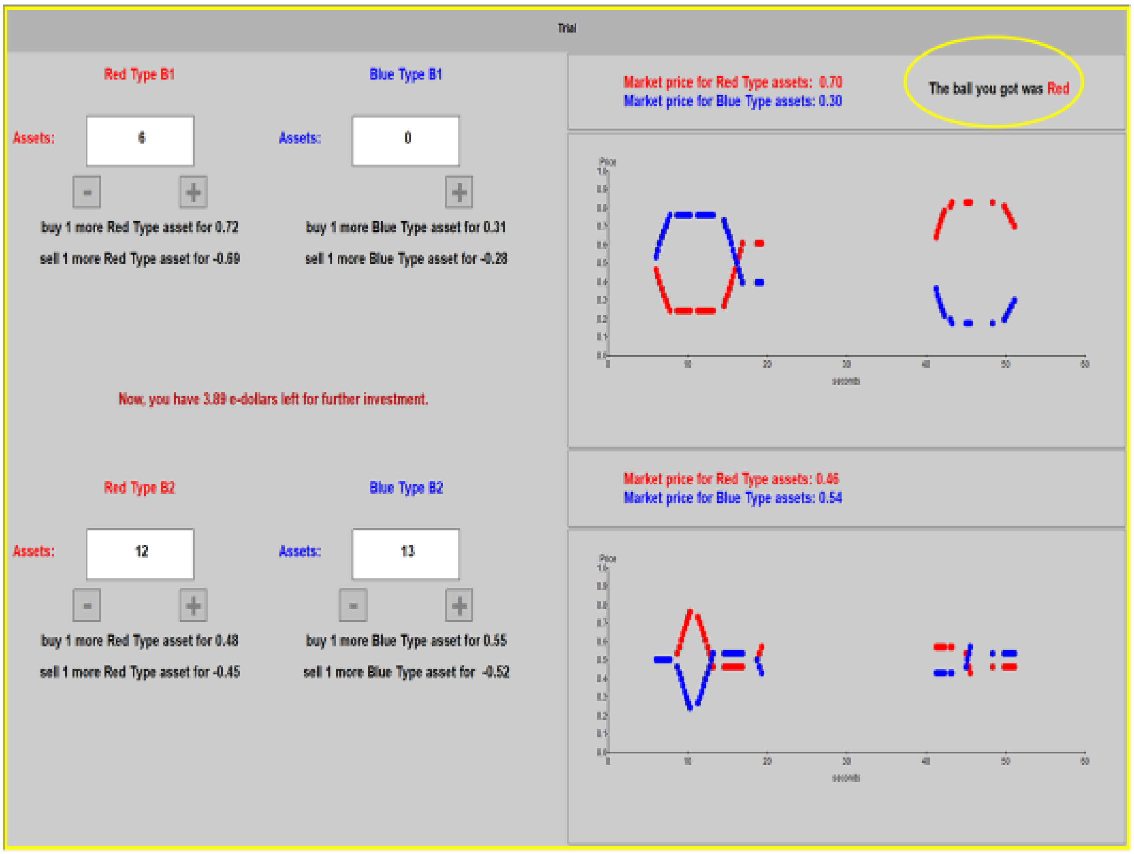
- You will have 20 e-dollars to trade/purchase assets in *both markets*.
- The 20 e-dollars will expire at the end of each round.
- Each round will last 1 minute. You have **8 rounds.**
- Practice rounds are **not paid** but you will get to see your payoffs at the end to give you an idea of how your payoffs will look like.



Trial
Red Type B1
Blue Type B1
Assets: 0
Assets: 0
?
Red Type Bucket
8 red balls and 2 blue balls
Blue Type Bucket
8 blue balls and 2 red balls
Market for Bucket 1
Red Type B2
Blue Type B2
Assets: 0
Assets: 0
?
Red Type Bucket
8 red balls and 2 blue balls
Blue Type Bucket
8 blue balls and 2 red balls
You have now understood the instructions for Stage 2.
Next, you will see the colour of the randomly selected ball from one of the buckets.
Press OK when you are ready.
Market for Bucket 2
OK



Trial
Red Type B1
Blue Type B1
Assets: 0
Assets: 0
?
The color of the private ball you received is:
RED
Red Type Bucket
8 red balls and 2 blue balls
Blue Type Bucket
8 blue balls and 2 red balls
Remember, you have 20 e-dollars to invest.
You have received a signal from Bucket 1
Red Type B2
Blue Type B2
Assets: 0
Assets: 0
You will now enter the market.
Press OK when you are ready.
OK



Decision Rule 1: **Prediction Market** *(two buckets-**PM2**)*

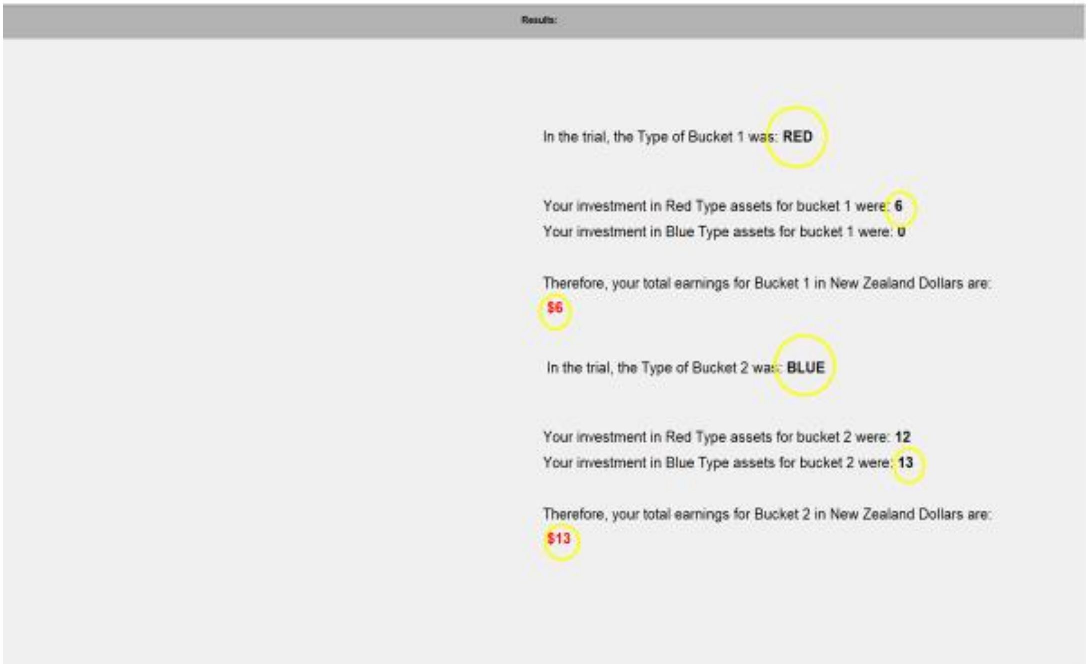
Stage 2-Payoff

After all trading rounds, the computer will **randomly** select **one round** and reveal the type of **both buckets** for that round.

- If the true type of the bucket turned out to be Red, you will be paid **1 e-dollar for every Red asset you purchased; 0 e-dollars for Blue.**
- If the true type of the bucket turned out to be Blue, you will be paid **1 e-dollar for every Blue asset you purchased; 0 e-dollars for Red.**

Your rewards for Stage 2, based on the decisions you made for the selected round, will be revealed to you at the end of today's experiment.

Payoff Example



Decision Rule 2: **Deterministic (DMD)**

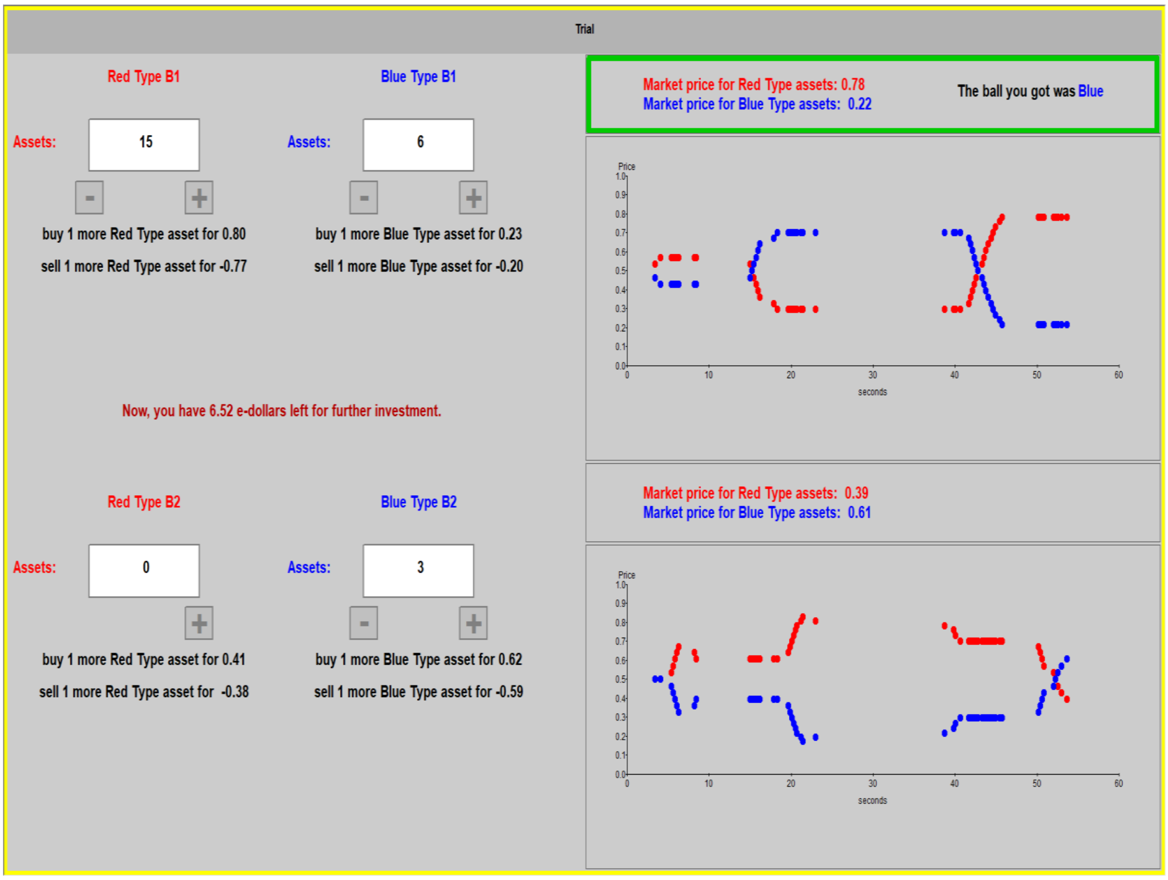
Stage 2-Payoff

After all trading rounds, the computer will:

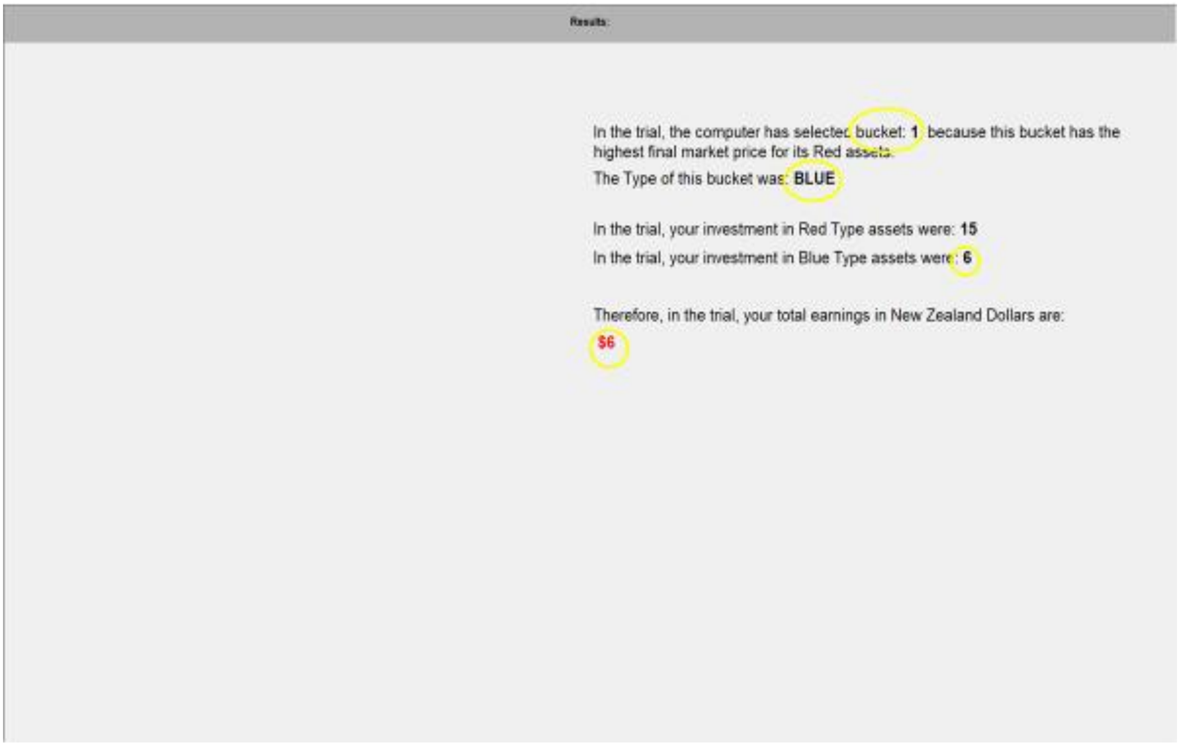
1. Select **one** round; randomly.
2. Select the bucket that has the ***highest final market price for its Red Type assets*** and;
3. Reveal the **true** type of the selected bucket.

You will be paid the owned assets which is consistent with the true type of the selected bucket; which is **1 e-dollar for every asset you purchased.**

Your rewards for Stage 2, based on the decisions you made for the selected round, will be revealed to you at the end of today's experiment.



Payoff Example



Decision Rule 3: **Stochastic (DMS)**

Stage 2-Payoff

After all trading rounds, the computer will:

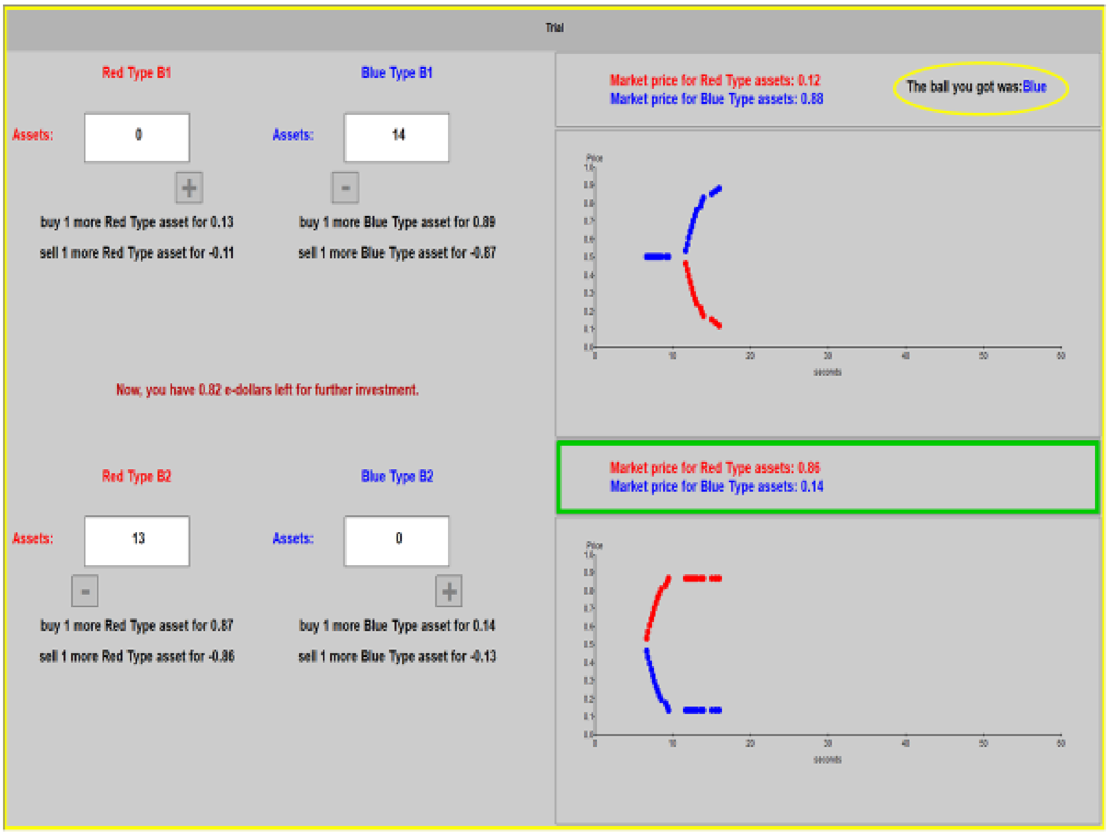
1. Select **one** round; randomly
2. One bucket will be chosen:

   – The bucket that has the ***highest final market price for its Red Type assets*** will be selected at a ***75% probability;***
   – The other bucket will be selected at a ***25% probability.***

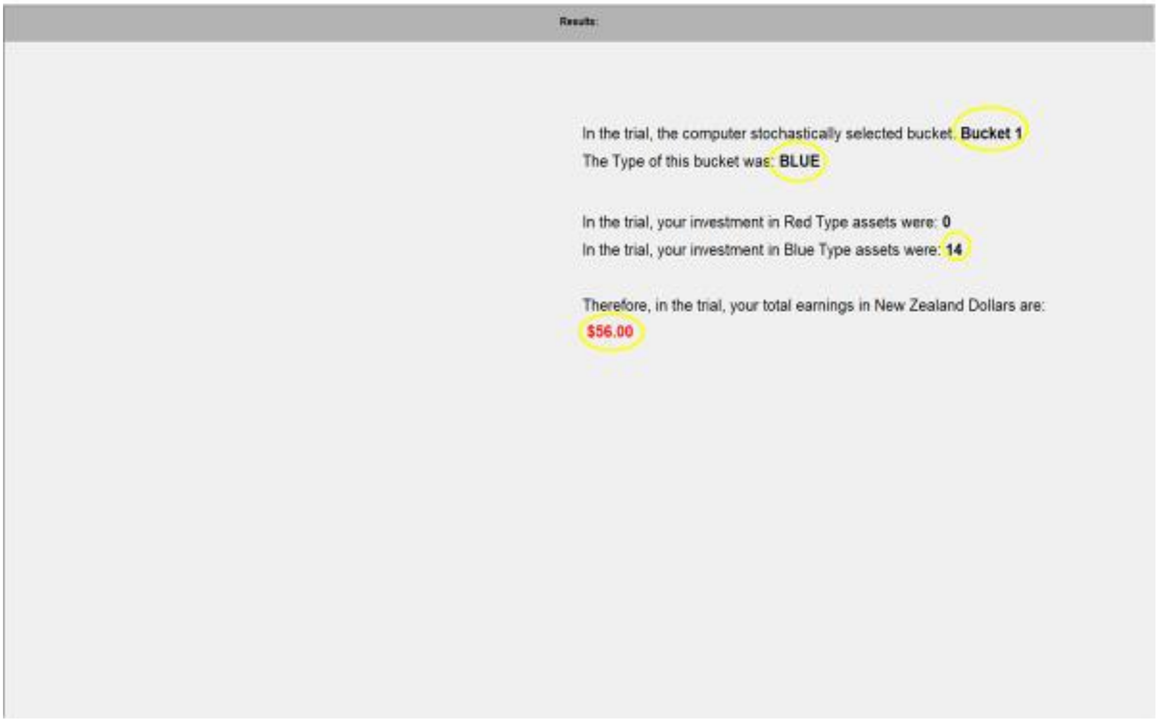
3. Reveal the **true** type of the selected bucket.

   – If the bucket is selected at 75% probability, your payoff will be increased by **4/3**
   – If the bucket is selected at 25% probability, your payoff will be increased by **4**

Your rewards for Stage 2, based on the decisions you made for the selected round, will be revealed to you at the end of today's experiment.



Payoff Example



Appendix D: Stage 3 Instructions

Stage 3

- You have now reached **Stage 3** of today's experiment where you are asked to complete a small lottery task and fill out a questionnaire.
- This will only take a few minutes.
- You will get an additional reward of **up to <u>$3.85</u>** for completing the lottery task.
- The answers you will provide to the questionnaire will have no impact on your overall rewards, which will only depend on the choices made in Stage 1 and Stage 2.

Appendix E: Demographic Questionnaire and Risk Lottery

Demographics Questionnaire

1) Please select the box that corresponds to your age

less than 18

18

19

20

21

22

23

24

25

26

27

28

29

30

More than 30

2) Gender (press the button that corresponds to your gender)

Male

Female

3) Which Ethnic group do you belong do?

New Zealand Maori

Pakeha/European

Pacific Islander

Asian

Middle Eastern

African

South American

Indian

Other

4) Please select your school

Business

Engineering

Sciences

Arts

Other

5) Have you taken a statistical course before?

Yes

No

Risk Lottery

Table: The Holt and Laury (2002) Multiple Price List (MPL) Approach.

The Multiple Price List Approach *a*` la Holt and Laury (2002)

Lottery ONE				Lottery TWO			
0.1	$2	0.9	$1.6	0.1	$3.85	0.9	$0.1
0.2	$2	0.8	$1.6	0.2	$3.85	0.8	$0.1
0.3	$2	0.7	$1.6	0.3	$3.85	0.7	$0.1
0.4	$2	0.6	$1.6	0.4	$3.85	0.6	$0.1
0.5	$2	0.5	$1.6	0.5	$3.85	0.5	$0.1
0.6	$2	0.4	$1.6	0.6	$3.85	0.4	$0.1
0.7	$2	0.3	$1.6	0.7	$3.85	0.3	$0.1
0.8	$2	0.2	$1.6	0.8	$3.85	0.2	$0.1
0.9	$2	0.1	$1.6	0.9	$3.85	0.1	$0.1
1	$2	0	$1.6	1	$3.85	0	$0.1

Appendix F

Correct aggregation

Let b represent the event that an urn is of Blue Type, r that an urn is of Red Type, R that a signal is a red ball and B that a signal is a blue ball. Where:

- $P(b)$ is the prior probability for an urn to be blue.
- $P(b|R)$ is the posterior after learning that a signal is a red ball
- $P(b|B)$ is the posterior after learning that a signal is a blue ball.

- $P(r)$ is the prior probability of the urn to be red.
- $P(r|R)$ is the posterior after learning that a signal is a red ball.
- $P(r|B)$ is the posterior after learning that a signal is a blue ball.

with the signal probabilities,

- $P(R|b)$ is the probability of drawing a red ball if the urn is blue
- $P(B|b)$ is the probability of drawing a blue ball if the urn is blue
- $P(R|r)$ is the probability of drawing a red ball if the urn is red
- $P(B|r)$ is the probability of drawing a blue ball if the urn is red

The posterior after learning the signal probabilities can be calculated using Bayes' Theorem,

$$P(b|R) = \frac{P(R|b) \cdot P(b)}{P(R)}$$

$$P(b|R) = \frac{P(R|b) \cdot P(b)}{P(R|b) \cdot P(b) + p(R|r) \cdot (1 - p(b))}$$

$$P(b|B) = \frac{P(B|b) \cdot P(b)}{P(B|b) \cdot P(b) + p(B|r) \cdot (1 - p(b))}$$

$$P(r|R) = \frac{P(R|r) \cdot P(r)}{P(R|r) \cdot P(r) + p(R|b) \cdot (1 - p(r))}$$

$$P(r|B) = \frac{P(B|r) \cdot P(r)}{P(B|r) \cdot P(r) + p(B|b) \cdot (1 - p(r))}$$

In the context of our experiment, at the beginning,

$$P(b) = P(r) = 0.5$$

$$P(R|b) = P(B|r) = 0.2$$

$$P(B|b) = P(R|r) = 0.8$$

Thus an agent who sees a red ball gets a posterior of,

$$P(b|\mathrm{R}) = \frac{0.2 \cdot 0.5}{0.2 \cdot 0.5 + 0.8 \cdot 0.5}$$

$$= 0.2$$

$$P(r|\mathrm{R}) = \frac{0.8 \cdot 0.5}{0.8 \cdot 0.5 + 0.2 \cdot 0.5}$$

$$= 0.8$$

and if the signal was a blue ball, a posterior of,

$$P(b|B) = \frac{0.8 \cdot 0.5}{0.8 \cdot 0.5 + 0.2 \cdot 0.5}$$

$$= 0.8$$

$$P(r|B) = \frac{0.2 \cdot 0.5}{0.2 \cdot 0.5 + 0.8 \cdot 0.5}$$

$$= 0.2$$

Aggregation will be assessed using a posterior of a hypothetical agent who knows both signals. Using updating for a second signal gives,

$$P(b|R,R) = \frac{P(R|b) \cdot P(b|R)}{P(R|b) \cdot P(b|R) + p(R|r) \cdot (1 - p(b|R))}$$

$$P(b|R,R) = \frac{0.2 \cdot 0.2}{0.2 \cdot 0.2 + 0.8 \cdot 0.8}$$

= 0.059

Therefore, after seeing the first red ball, an agent knows that the probability for the blue urn decreases from 0.5 to 0.2 and drawing a second ball decreases the probability from 0.2 to 0.059. If our first signal is a blue ball and second signal is a red ball then our probability is,

$$P(b|B,R) = \frac{P(B|b) \cdot P(b|R)}{P(B|b) \cdot P(b|R) + p(B|r) \cdot (1 - p(b|R))}$$

$$P(b|B,R) = \frac{0.8 \cdot 0.2}{0.8 \cdot 0.2 + 0.2 \cdot 0.8}$$

= 0.5

In our experiment, all possible 2-ball posteriors are given as,

$P(b|R,R) = 0.059$ $\qquad$ $P(r|R,R) = 0.941$

$P(b|B,B) = 0.941$ $\qquad$ $P(r|B,B) = 0.059$

$P(b|B,R) = 0.5$ $\qquad$ $P(r|B,R) = 0.5$

We quantify the error by the difference between the market price we observe and the perfect aggregation probability. We use error using log-transformed probabilities,

$$err_i = \left(log\, \frac{P^M{}_i}{1-P^M{}_i} - log\, \frac{P^B{}_i}{1-P^B{}_i}\right)^2$$

where perfect aggregation is our market for urn i being equal to the posterior $P^{B}{}_{i}$ of a hypothetical agent who knows both signals. We use logit transformation as an error estimator instead of absolute differences because we assume that our probabilities can be extreme (move towards 0.1).

Appendix G

Risk Aversion Classifications Based on Lottery Choices.

Table G.1: The Holt and Laury procedure (2002) for the Multiple price list (MPL) method.

Number of Safe Choices	Range of Relative Risk Aversion for $U(x) = x^{1-r}/(1-r)$	Risk Preference Classification
0-1	$r < -0.95$	highly risk loving
2	$-0.95 < r < -0.49$	very risk loving
3	$-0.49 < r < -0.15$	risk loving
4	$-0.15 < r < 0.15$	risk neutral
5	$0.15 < r < 0.41$	slightly risk averse
6	$0.41 < r < 0.68$	risk averse
7	$0.68 < r < 0.97$	very risk averse
8	$0.97 < r < 1.37$	highly risk averse
9-10	$1.37 < r$	stay in bed

Source: Holt and Laury (2002)

1) Holt and Laury's (2002) ''Number of Safe Choices'': The lottery task has a list of 10 decisions (see Table G.1). Each decision is a paired choice, between option A (safe) and option B (risky). Participants must choose between the safe options A and the more risky options B. The participant's risk attitude is defined by the number of A-choices: 1–3 (risk-loving), 4 (risk neutral), and 5–10 (risk-averse).

2) Eckel et al.’s (2010) ‘’inconsistency’’: an inconsistent participant is one that, as she moves down the lottery tasks, switches back to the safe option (A) after having chosen a risky option (B).