In this thesis, I develop new elicitation methods for people's preferences and private information and experimentally test them. Combined with standard tools in neoclassical economics, these new elicitation methods disentangle the confounding motives and systematic biases behind people's preferences, beliefs, and information processing. In particular, Chapter 2 recovers individual preferences over the trade-off between the aggregate wealth and the distributional equity behind the veil of ignorance and investigates their relationships with risk preferences. Chapter 3 elicits individual preferences over expert and quack tests and identifies how failures of contingent reasoning contribute to choices of quacks. Chapter 4 tests the validity of the Bayesian market in eliciting private information and examines the role of belief disturbances. Chapter 5 relaxes standard assumptions and proposes two simple betting mechanisms to extract private information. This thesis improves our understanding of what people like and what people think in several new contexts.

Yan Xu holds an MPhil in Economics from the Tinbergen Institute (2016). After that, she joined the Erasmus School of Economics at the Erasmus University Rotterdam as a Ph.D. candidate, under the supervision of Prof. Aurélien Baillon and Prof. Dražen Prelec. Her research interests span behavioral economics, experimental economics, and decision theory. She currently works as an assistant professor in economics at the University of Vienna.

Eliciting Preferences and Private Information

Yan Xu

768

tinbergen
institute

Yan Xu

Eliciting Preferences and Private Information:

Tell Me What You Like and What You Think

Erasmus University Rotterdam

Eliciting Preferences and Private Information:
Tell Me What You Like and What You Think

ISBN 978 90 3610 624 5
Cover design: Crasborn Graphic Designers bno, Valkenburg a.d. Geul

This book is no. **768** of the Tinbergen Institute Research Series, established through cooperation between Rozenberg Publishers and the Tinbergen Institute. A list of books which already appeared in the series can be found in the back.

Eliciting Preferences and Private Information: Tell Me What You Like and What You Think

Het uitlokken van voorkeuren en privéinformatie:
Vertel me wat je leuk vindt en wat je denkt

Thesis

to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus

prof.dr. R.C.M.E. Engels

and in accordance with the decision of the Doctorate Board.
The public defence shall be held on
Thursday October 22, 2020 at 09:30 hours

by

Yan Xu

born in Changzhou, China

Erasmus University Rotterdam

Doctoral Committee:

Promotors:	prof.dr. A. Baillon
	prof.dr. D. Prelec
Other members:	prof.dr. H. Bleichrodt
	dr. C. Li
	dr. E. Tsakas

Acknowledgement

I would like to express my sincere gratitude to my supervisor Aurélien Baillon. I would never complete my Ph.D. without your careful and patient guidance. In the past four years, Aurélien provided me tremendous support on both my academic career and my everyday life. He respected my ideas and encouraged me to explore research questions I am genuinely interested in. Whenever I was stuck with my research progress or went through the life crisis, he was always there, being understanding, sharing experiences, and providing wise suggestions.

I am also grateful for my co-supervisor Dražen Prelec. Thanks for hosting my academic visit at the MIT Sloan School of Management. It was a valuable experience. I enriched my research perspectives. His insightful comments improved Chapter 5. I am particularly benefited from his lectures on Psychology and Economics. I learned from Dražen the important ingredients of developing good research ideas: being curious and asking the most basic questions. Chapter 3 of the dissertation was inspired by a discussion in his class.

I would like to thank Peter Wakker for being a role model. Max Weber probed the question of science as a vocation over one hundred years ago. I never fully understood it, and I questioned the meaning of my research and the career prospects a lot at the beginning of my Ph.D. But over the years, I gained a better understanding of the question through Peter's enthusiasm for research. I was continuously encouraged by his office light on Friday night and regular updates of the annotated bibliography and determined to pursue an academic career.

I am lucky to be a member of the Behavioral Economics research group of Erasmus University Rotterdam. I thank all my colleagues and friends for their stimulating discussions in group meetings, valuable feedback on my ideas, presentations, and experiments. In no particular order, I thank Han Bleichrodt, Kirsten Rohde, Jan Stoop, Chen Li, Vitalie Spinu, Georg Granic, Tong Wang, Jingni Yang, Paul van Bruggen, Benjamin Tereick, Cem Peker, Xiao Yu, Merel van Hulsen, Francesco Capozza, Marine Hainguerlot, Sophie van der Zee, Aysil Emirmahmutoglu, Hendrik Rommeswinkel, IIke Aydogan, Yu Gao, Ning Liu, and Zhenxing Huang. Special thanks to Jan Heufer for the great collaboration we had.

I would like to thank my coauthor Jason Shachat. He was my mentor when I worked at the FEEL lab at Xiamen University eight years ago. At that time, I did not have particular interests in my major in Finance and had no idea about what I should do in the future. Jason introduced me to experimental economics and showed me how to conduct research from the beginning. Thanks for having confidence in me and being encouraging all the time. I am also grateful for my former colleagues at the FEEL lab. Lijia Wei, Lijia Tan, Weiwei Zheng, Yun Wang, Sen Geng, Xiaoyi Gao. I always feel magical to meet you guys at different conferences.

I am very grateful to my close friends. Thank you Yu Zhang and Ye Tao for always encouraging and supportive. We've been friends for over half of my life, and I am lucky to have your companies on the journey. Thanks my friends at the Tinbergen Institute for their help on my research and their company outside the academics (Lingwei Kong, Zichen Deng, Junze Sun, Huaiping Yuan, Xun Gong, Jing Chen, Mengheng Li, Rui Zhuo, Shihao Yu, Yuan Yue, Yuhao Zhu, Zhen Li). I enjoyed the time together with you on traveling, cooking, games, and restaurant hunting.

Finally, I would like to thank my parents and my brother for their unconditional support and love. You always respect my decisions and being proud of every small step I made. I am also grateful to Dr. F. Zhu. You were on my side and walked me through all difficult and joyful times.

I also place on record, my gratitude to one and all, who directly or indirectly, have lent their hand in this venture. All errors and omissions in this dissertation are mine.

Yan Xu
Rotterdam, June 2020

Contents

1 Introduction 1

1.1 Tell me what you like: eliciting preferences by constructing trade-off problems . . . 2

1.2 Tell me what you think: eliciting private information by aligning incentives 4

2 Measuring tastes for equity and aggregate wealth behind the veil of ignorance 7

2.1 Introduction . . . 7

2.2 Literature review . . . 12

2.3 Experimental design . . . 15

2.4 Reduced form analysis . . . 17

2.4.1 Aggregate and individual treatment effects . . . 18

2.4.2 Reduced form demand estimation . . . 21

2.5 Nonparametric analysis . . . 23

2.5.1 Testing rationality and homotheticity . . . 23

2.5.2 Rationality results and comparisons with other studies . . . 25

2.5.3 The relative convexity of Risk-self versus VoI-self preferences . . 30

2.5.4 Social preference classification stability . . . 34

2.6 Structural equation estimation and analysis . . . 36

2.6.1 Subjective expected utility specification and estimation approach 36

2.6.2 Individual level parametric estimation . . . 38

2.6.3 Comparison with other experiments . . . 40

2.7 Conclusion . . . 43

Appendix 2.A Experimental instructions . . . 44

2.A.1 Instructions for the first stage . . . 44

2.A.2 Instructions for the second stage . . . 47

Appendix 2.B Type classifications based on alternative criteria . . . 50

Appendix 2.C Additional materials for parametric analysis . . . 51

3 Revealed preferences over experts and quacks and failures of contingent reasoning 55

3.1 Introduction . . . 56

3.2 Theoretic framework for expert and quack tests . . . 60

3.2.1 Setup: states, signals, and tests . . . 60

3.2.2 Distinguishing experts and quacks in a rational benchmark . . . 61

3.2.3 Eliciting preference over tests via linear budget sets . . . 63

3.3 Experimental design . . . 65

3.3.1 Tasks . . . 65

3.3.2 Procedures . . . 68

3.4 Experimental results . . . 69

3.4.1 The failure in distinguishing experts and quacks . . . 70

3.4.2 The failure in evaluating experts and quacks . . . 71

3.5 Unexplained mechanisms . . . 73

3.5.1 Belief updating bias . . . 73

3.5.2 Sub-optimal bet choices . . . 79

3.5.3 Intrinsic preferences over test characteristics . . . 80

3.6 Explained mechanisms: contingent reasoning and decision rules . . . 82

3.7 Discussion and implication . . . 88

3.8 Conclusion . . . 90

Appendix 3.A Proofs . . . 91

Appendix 3.B More details on the experimental design . . . 92

Appendix 3.C Additional analyses . . . 94

3.C.1 Additional analyses on belief updating bias . . . 94

3.C.2 Additional analyses on sub-optimal bet choices and test characteristics . . . 94

3.C.3 Additional analyses on demographics, attitudes, and cognitive abilities . . . 100

Appendix 3.D Experimental instructions and screenshots . . . 102

4 Will Bayesian markets induce truth-telling? —An experimental study 109

4.1 Introduction . . . 109

4.2 Bayesian market mechanism . . . 113

4.2.1 Model setup . . . 113

4.2.2 Theoretical predictions of Bayesian markets . . . 115

4.3 Experimental design and procedures . . . 116

4.3.1 Market structure . . . 116

4.3.2 Experimental treatments 118
4.3.3 Experimental procedures 120
4.4 Experimental results 121
4.4.1 Aggregate truthfulness rate 121
4.4.2 Individual truthfulness rate 122
4.4.3 Heterogeneity in strategy evolution 123
4.4.4 Heterogeneity in signal effects 123
4.5 Explaining treatment effects: bid and ask prices 125
4.5.1 Aggregate bid and ask prices 126
4.5.2 Shirking, speculation, and updating bias 128
4.5.3 Heuristics 131
4.6 Conclusion and discussions 134
Appendix 4.A Parameter settings for the experiment 136
Appendix 4.B Experimental instructions (1HA) 137
Appendix 4.C Additional analyses on the bid/ask prices in 8HA treatment . 143

5 Simple bets to elicit private signals 147
5.1 Introduction 147
5.2 Betting on exogenous ratings 150
5.2.1 Signals, ratings, and beliefs 150
5.2.2 The bets 153
5.3 Betting on endogenous ratings 156
5.3.1 Agents, their signals, and their beliefs 156
5.3.2 The games 157
5.4 Discussion 160
5.4.1 Limitations and related literature 160
5.4.2 Practical implementation and examples 163
5.5 Conclusion 164
Appendix 5.A Proofs for the single-agent setting 165
Appendix 5.B Proofs for the game setting 169

6 Conclusion 171

Summary 174

Samenvatting (Summary in Dutch) 176

Bibliography 179

Chapter 1

Introduction

In economics, individuals' preferences and private information are primitives for their decisions. The classical model describes human choices under uncertainty as individuals maximizing a combination of a belief component (which depends on what they think) and a utility component (which describes what they like). Consider the following prototypical economist's conception of human behavior, adapted from Rabin (1998), Rabin (2002), and DellaVigna (2009). A decision-maker (she) faces a choice set $\mathcal{X}$ and chooses the one that maximizes her expected utility:

$$\max_{x \in \mathcal{X}} \sum_{s \in \mathcal{S}} p\,(s \mid t)\, U\,(x \mid s)\,.$$

The uncertainty is described as probabilistic states of the world $\mathcal{S}$. The utility function $U(x \mid s)$ is defined over the DM's payoffs in the state s. The decision-maker also has the private information t, and she updates and formulates a belief distribution $p(s \mid t)$ over the states.

The standard model imposes many implicit assumptions on human nature. Individuals only care about their own payoffs; their preferences are well-behaved and independent of the framing of the choice problem; they process private information appropriately and update beliefs as Bayesian agents; they are fully attended, immune from emotions, and capable of solving the maximization problem. Many laboratory experiments and empirical evidence show individuals' preferences, beliefs, and decision-making processes are non-standard.[1]

[1] Related reviews include Kahneman, Knetsch and Thaler (1991), Camerer and Thaler (1995), Starmer (2000), Mullainathan and Thaler (2000), Rabin and Thaler (2001), Charness and Rabin (2002), Kahneman (2003*b*), Levitt and List (2007), Moore and Healy (2008), Gigerenzer, Hertwig and Pachur (2011), Barberis (2013), Thaler and Ganser (2015), Lerner et al. (2015), DellaVigna (2009) ect. The social preference $u(x_i, x_o \mid s)$ and the reference dependent preference $U(x \mid s, r)$ are two examples of non-standard preferences.

In many settings, the deviations from standard assumptions do not hinder economic research. For instance, the assumptions on preferences and beliefs are necessary axioms or simplifications of the stylized facts that people are self-interested and well-informed at the aggregate level. The deviations are often not negligible in individual decision-making scenarios when the research question is directly on what people like and what people think. Over the past four decades, behavioral economics focuses on such settings. To explain non-standard preferences, beliefs, and decision-making processes, it integrates more realistic notions of human psychology into the neoclassical economics and formulates new models and empirical tests of human decisions (see Camerer and Loewenstein (2003), Camerer (1999), Kahneman (2003*a*)).

The deviations and the resulting behavioral biases impose new requirements and challenges on the elicitation of preferences and private information. This thesis develops *clean* and *robust* methods and elicit what people like (individual preference) and what people think (private information) in several decision scenarios. The criterion "clean" requires the method to isolate different confounding motives behind people's choices. The criteria "robust" requires that the elicited preferences are stable against noises due to trembling hands or cognitive disturbances. In the following subsections, I will introduce the elicitation methods for preferences and private information separately. The preference elicitation is based on the classical revealed preference approach, and the private signal is elicited by creating incentives for truth-telling.

1.1 Tell me what you like: eliciting preferences by constructing trade-off problems

In the first economics class, we learn that individuals' preferences over good x and y are described by indifference curves. A well-behaved preference is shaped by a convex downward sloping indifference curve, reflecting the intrinsic trade-offs between two goods. Getting one more unit of apples (good x) is at the cost of forgoing some units of oranges (good y). We can recover a DM's preference if we ask her to make trade-offs between different amounts of apples and oranges. From the revealed choice perspective, it is equivalent to providing the decision-maker several budget sets and then, recovering her indifference curves from her consumption bundles. Unlike apples and oranges,

Non-standard beliefs include overconfidence and belief updating biases, such as confirmation bias, base-rate neglect, and under(over)-inference. Examples of non-standard decision processes include limited attention and memory (decide based on a subset of $\mathcal{X}$), bounded rationality (cannot solve the maximization problem), and framing effects (have different solutions for the the maximization problem).

the goods in many settings are non-standard, and the revealed preference approach is not directly applicable. I re-construct trade-off problems and elicit preferences in new contexts. Chapter 2 measures individual preferences over lotteries, economic equity, and aggregate wealth. Chapter 3 recovers preferences over useful and useless tests (also called information structures or information sources).

In Chapter 2 titled *"Measuring tastes for equity and aggregate wealth behind the veil of ignorance"* (joint with Jan Heufer and Jason Shachat), we propose an instrument to answer one of the oldest yet controversial questions in economics — how do people make trade-offs between the aggregate wealth and the distributional equality among social members? Do people prefer a policy that generates overall prosperity but delivers insufficient equity or vice versa? We exploit the potential of Harsanyi's Veil of Ignorance (VoI) as a tool and create a novel experiment to measure such preferences at the individual level. In particular, a decision-maker faces choice problems varying in the relative "expensiveness" of efficiency and equity and decides the optimal monetary allocations for a two-person economy without knowing her status of being rich or poor. Each choice problem contains a rich set of trade-offs, making sure the elicited preferences are robust against trembling hands in individuals' choices.

However, the intrinsic uncertainty in the VoI framework conflates individuals' distributional preferences with their risk preferences. The choice of an equal allocation for the wealth distribution problem does not necessarily mean the decision-maker cares about equity; she might be just risk-averse. We resolve this challenge by pairing each wealth distribution problem with a risk portfolio choice problem. Both problems share a common budget set and the same distribution over an individual's own wealth. The risk portfolio choice provides a clean identification for the distributional preference. Individuals' choices in risk problems serve as a benchmark to evaluate whether their wealth distribution choices exhibit equity or efficiency preferring tastes. The paired choice problems also allow us to investigate how different motives interact with each other and how they jointly determine choices.

Chapter 3, titled *"Revealed preference over experts and quacks and failures of contingent reasoning"*, elicits individual preferences over tests and investigates how people formulate the judgment over the usefulness of tests. It studies the scenario wherein people face incomplete information about the payoff-relevant states of the world and resort to tests (e.g., analysts, medical diagnoses, or psychic octopuses) to obtain information. Each test represents an information source and will generate signals that are informative to the true state. This fundamental difference imposes new requirements for information processing. When there are many tests available, how do people evaluate and choose tests? Are they able to avoid useless quacks and identify genuinely useful experts? Are they over-paying for quacks and under-paying for experts, and why?

I formalize the theoretical framework for expert and quack tests. The usefulness of a test, defined as the expected benefit for the decision problem from having or not having the test, is a joint output of the decision problem at hand, individuals' preferences and belief formations, and their interactions reflected in the reasoning process. In particular, individuals choose tests before they actually acquire a piece of information. Such ex-ante evaluations require that individuals anticipate all possible signals, correctly formulate posterior beliefs for each signal and thus best-respond, and learn how different structures of signals influence the decision problem. This evaluation process identifies the mechanisms behind people's choices of tests.

Using a novel linear budget experiment, I elicit people's preferences over tests and examine the underlying mechanisms. Individuals face a rich and structured choice set of expert and quack tests and choose their favorite ones through a graphic coloring task. In particular, the choice problem is equivalent to making trade-offs between receiving good news in one state versus in the other on a linear budget. The elicited consumption bundles of two state-specific accuracies reflect individuals' preferences over the usefulness, Blackwell informativeness, and skewness for a set of ”affordable” tests. I construct fourteen budgets with multiple prices and expenditure levels. They generate rich variations both within and across the choice set of tests, thus providing diagnoses for different behavioral biases and decision patterns.

1.2 Tell me what you think: eliciting private information by aligning incentives

Private information is essential for us to gather knowledge and further guide better decision-making. Researchers run social-economic surveys on people's opinions and attitudes. Review sites like Rotten Tomatoes or Yelp collect our tastes for movies and experiences with restaurants. The elicitation of private information takes another approach. It relies on how to create monetary incentives to encourage truth-telling from respondents. For instance, we can promise a decision-maker a monetary compensation scheme, under which truthfully reporting her private information yields a higher expected payoff than untruthful reports. In many cases, private signals are subjective and unverifiable, making it challenging to align incentives. How can we assure that respondents' answers are attentive and truthful when the "ground truth" is not available for the assessment?

I start to answer this question via an experimental test of a market-based truth-telling elicitation mechanism. Chapter 4, titled *"Will Bayesian markets induce truth-telling?—An experimental study"*, tests the performance of Bayesian markets in inducing private signals and investigates how the performance is affected by individuals' belief systems. In a Bayesian market, each participant has a chance to trade an asset whose value is determined by other agents' trade decisions. According to Bayesian reasoning, individuals with different private signals evaluate the value of the asset differently, making the transaction possible in the market. Individuals will reveal their private signals through their choices of buying or selling a belief asset. The performance of Bayesian markets hinges on the concept of Bayesian Nash equilibrium, with which truth-telling is preferred when the decision-maker believes other participants are also truth-telling. I construct a laboratory Bayesian market to trade such belief assets and manipulate individuals' beliefs over others' truthfulness through different group compositions of human agents and algorithm agents.

I find Bayesian markets effectively induce truthful revelations when participants believe that others are truthful. A majority of subjects submit their private signals and form correct posterior valuations of the asset. When participants suspect that some of their opponents may lie, Bayesian markets become less effective, and bubbles arise in the market. The mechanism of how bubbles are formed explains the impact of belief noises on the validity of Bayesian markets. Due to speculative buyers in the market, participants are more likely to under-infer their private signals. They over-predict the value of the asset and thus are more likely to buy the asset. Such over-buying inclinations raise the ex-post realization of the asset value. In the end, people ignore their private signals and chase trends in the market, leading to market bubbles.

The findings in the Bayesian market motivate us to design new elicitation methods for private signals. We need mechanisms that are less dependent on belief assumptions and easy to implement in practice. In Chapter 5, titled *Simple bets to elicit private signals* (joint with Aurelien Baillon), we introduce two simple betting mechanisms, Top-Flop and Threshold betting, to elicit people's unverifiable signals. We deviate the standard elicitation methods and explore the multi-dimension of survey questions. There is a collection of similar items (questions), and each item has a rating. A decision-maker receives a binary private signal about one item and bets on its rating relative to that of another item in the collection. For instance, in the Top-Flop betting with a collection of similar movies, the decision-maker bets on or against whether the movie she just watched having a higher score than another, random movie. In the Threshold betting, she bets which one of two movies will exceed a threshold score. Both mechanisms have transparent payment rules and are simple to implement for many scenarios regarding people's tastes and experiences. We further establish the theoretical conditions ensuring that Top-Flop and Threshold betting properly reveal individuals' private signals through their chosen bets. Compare to other elicitation mechanisms, our methods relax the assumptions on common prior and are robust against risk attitudes.

Chapter 2

Measuring tastes for equity and aggregate wealth behind the veil of ignorance[1]

Abstract: This chapter proposes an instrument to measure individuals' social preferences regarding equity and efficiency behind a veil of ignorance. We pair portfolio and wealth distribution choice problems which have a common budget set. For a given bundle, the distribution over an individual's wealth is the same for both problems. The portfolio choice serves as a benchmark to evaluate whether the wealth distribution choice exhibits equity or efficiency preferring tastes. We report experiments using a within-subject design testing the veracity of this instrument. We find clusters of equity preferring, efficiency preferring, and egoist individuals through reduced form, revealed preference, and structural estimation analyses. We further find reduced form demand functions for risky assets are independent of social preference type classification.

2.1 Introduction

Economic policies often strive to balance maximizing total wealth, what we call efficiency, and delivering sufficiently equal individual shares of that wealth, what we call equity. Failing too miserably in either generating overall prosperity (Acemoglu, Johnson and Robinson, 2002; Rodrik and Wacziarg, 2005; Joseph, 2008) or delivering sufficiently equitable prosperity (Alesina and Rodrik, 1994; Alesina and Perotti, 1996;

[1] This chapter is based on Heufer, Shachat and Xu (2019).

Benabou, 2000) can lead to political and social instability. Knowledge of individuals' preferences over efficiency-equity trade-offs is key to performing this balancing act.

In this paper, we present a new instrument to measure these individual efficiency-equity preferences and demonstrate its efficacy. Researchers have taken multiple tacks to identify and measure these preferences. Our approach measures preferences behind a veil of ignorance, VoI hereafter, and disentangles pure efficiency-equity preferences from those of risk aversion over one's own uncertain wealth. To avoid the ambiguity concerns found in Rawls's (1958) VoI formulation, we adopt a formulation in which each potential placement in society is equally likely (Harsanyi, 1953, 1976). We take a standard economics approach to uncovering individual preferences. If a decision maker, DM hereafter, is a price taker and technology or expenditure constrain her set of feasible choices, then her choice equates her willingness to trade-off one choice element for another with their relative prices. We incorporate this idea by having the DM select her preferred alternative from a linear budget set.

At the core of our instrument is a pair of standard consumer choice problems. Specifically, in each problem the DM chooses a two element commodity bundle, $(x_i, y_i) \in \mathbb{R}^2_+$, where i is the problem type, from a budget set $qx_i + y_i = z$ with relative price $q \geq 1$ and expenditure $z > 0$. In the first type of the problem, VoI, the DM is a member of a two-person economy and chooses a pair of wealth levels, one each for the Poor (x_{VoI}) and Rich (y_{VoI}) individuals. When making this choice, the DM knows there is a fifty percent chance she will be Poor and the other individual Rich and a fifty percent chance it will be vice versa. In this case, the linear budget set is a wealth profile possibilities frontier and the "price" is the constant loss rate in efficiency from redistributing a unit of wealth from the Rich to the Poor position. When this price is strictly greater than one, total wealth is maximized by assigning all wealth to the Rich and none to the Poor position. To maintain the roles of Rich and Poor we further restrict the DM to choosing profiles lying above the forty-five degree line, the locus of equal wealth profiles. Figure 2.1a depicts the commodity space under this restriction. Figure 2.1b depicts an example of the DM's choice, $(x^*_{\text{VoI}}, y^*_{\text{VoI}})$; the tangent indifference curve we draw is not necessary but helps us to set the stage to ask questions regarding her rationality. But for now, using only this choice, what can we conclude about the DM's preferences regarding efficiency and equity?

We posit not much. This choice does not reflect her pure preferences in terms of the trade-off between aggregate wealth levels and its distribution because it is conflated with the DM's risk preferences over her own terminal wealth. We introduce a second choice problem, Risk, to provide control for such risk preferences. In the Risk problem,

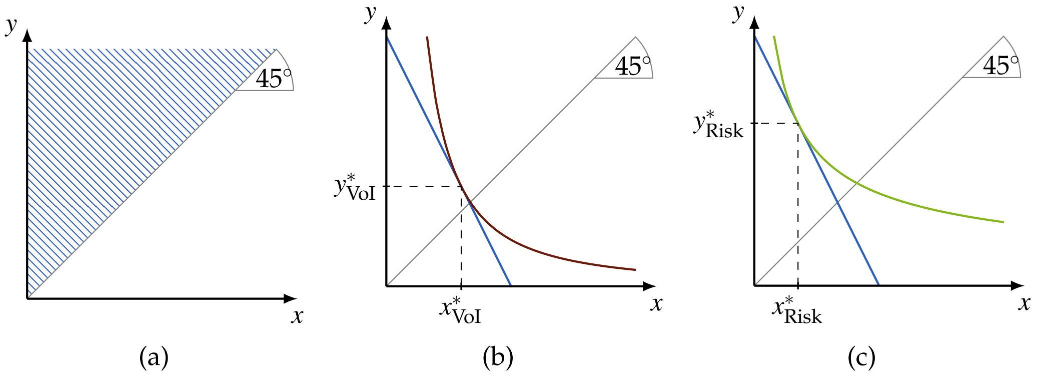


Figure 2.1: The commodity space.

the DM selects a state-contingent wealth profile from a set of feasible portfolios. These portfolios generate exactly the same distribution over individual wealth as the one from the VoI problem.

In a Risk problem, the DM chooses a portfolio of contingent claim assets in a world with two equally likely states, Good and Bad.[2] We restrict the DM's portfolio $(x_{\text{Risk}}, y_{\text{Risk}})$ to satisfy $y_{\text{Risk}} \geq x_{\text{Risk}} \geq 0$, thus preserving the roles of Good and Bad for the two states. The commodity space is the same as in the VoI problem. As before, the DM chooses a portfolio from a linear budget constraint $qx_{\text{Risk}} + y_{\text{Risk}} = z$ where $q \geq 1$; the price of insurance is at best actuarially fair. Figure 2.1c depicts an example where the consumer makes a choice facing the same price and expenditure as the VoI problem in Figure 2.1b.

How do these two problems differ? In the VoI problem, the DM receives one of the two possible amounts and somebody else the other, whereas in the Risk problem, the DM receives one of the two possible amounts and there is nobody else. How are these problems similar? For a fixed price and expenditure, a bundle on the VoI budget line generates the same marginal distribution over the DM's terminal wealth as the same bundle on the Risk budget line. Hence, the choice in the Risk problem provides us a benchmark for the VoI choice problem that controls for the DM's risk preference.[3]

[2] In our experiments we use the labels High and Low reward for both the VoI and Risk problems. For exposition in this paper, we use the corresponding labels Rich and Poor for VoI problems and Good and Bad for Risk problems. We also use the terms High and Low when we speak generically of a choice problem.

[3] One could argue that the DM's VoI and Risk choices differ only due the DM's custodial duties managing the other person's risk rather than any equity or efficiency concerns. A number of recent studies suggest that this is not the case. These studies compare the degree of risk aversions exhibited when a DM makes a series of binary lottery choices for themselves and the same series of lottery choices where they and another individual are beneficiaries; we do note the same lottery realization is used to determine the rewards, creating perfect positive rather than negative correlation. When lotteries are solely in the gain domain the majority of studies, such as Andersson et al. (2014); Vieider et al. (2016);

We define three classifications of social preferences for a DM's choice by comparing her chosen pair of quantities demanded $x_{\mathrm{VoI}}(q,z)$ and $x_{\mathrm{Risk}}(q,z)$ for a budget with price ratio q and expenditure z. First, when these quantities are the same, then the DM is not bothered by the societal implications of her choice and we call her an *egoist*. Second, if the DM gives more to the Poor in the VoI problem than to the Bad in the Risk problem, $x_{\mathrm{VoI}}(q,z) > x_{\mathrm{Risk}}(q,z)$, we call her *equity preferring*. Third, if the DM gives a smaller amount to the Poor, $x_{\mathrm{VoI}}(q,z) < x_{\mathrm{Risk}}(q,z)$, we call her *efficiency preferring*. In the example we provide in Figure 2.1 the DM prefers equity for the given budget line.

We complete our instrument by compiling forty rounds of such paired problems, where each round differs in the price and the expenditure level. Individuals make choices graphically by moving a computerized slider to select proportions of an amount of money between High and Low rewards.[4] We evaluate the efficacy of our instrument by conducting financially incentivized experiments with 92 students from Xiamen University. Our instrument necessitates a within-subject experimental design in which each participant makes both VoI and Risk choices. The within-subject design paired with our analytic approaches enable us identify of individual heterogeneity in social preference types. This heterogeneity can be misdiagnosed as homogeneous egoist behaviour in population level and between subject experimental studies.

In our first analysis of the data, we evaluate in which of the three social preference categories the average difference between the amounts given to Low rewards in the VoI and Risk problems lies. When examining these differences for all paired choices over all participants we find a very narrow confidence interval around zero. Thus, if we assume homogeneity across individuals then we conclude the population is egoist. However, inspection of each individual's average difference in the forty paired choices reveals large sized clusters for each social preference class. For each of three clusters we estimate reduced demand functions for both the allocation to Poor, $x_{\mathrm{VoI}}(q,z)$, and Bad, $x_{\mathrm{Risk}}(q,z)$. Surprisingly, a Chow-test does not reject that the three clusters have a common reduced demand function for x_{Risk}. The classification of three social preferences is not correlated with risk preferences.

Our second analysis of the data utilizes revealed preference arguments, not requiring the strong statistical assumptions of our first analysis, which allow inferences regarding the deeper structure of preferences and the rationality of individuals. First, we find that participants' choices exceed typical standards of adherence to the Generalized Axiom of

Füllbrunn and Luhan (2017, 2015), find no difference in risk aversion. Two notable exceptions are Sujoy et al. (2011); Pahlke, Strasser and Vieider (2015).

[4] We feel this presents a more intuitive approach to participants than choosing a point on a budget line, a popular frame of budget choice experiments since Choi et al. (2007*b*).

Revealed Preference, GARP hereafter. This implies that most individuals' behaviours, for both their Risk- and VoI-selves, are consistent with the maximization of a concave utility function. Second, we find that a large proportion of our participants do not make choices consistent with the maximization of a homothetic concave utility function, nullifying an ex-ante concern that our design makes homothetic consistent choices focal.

Finally, we make intra-personal comparisons of each individual's VoI and Risk domain utility functions. In this exercise we find that about twenty-eight percent of the subjects' Risk utility functions are globally more concave than their VoI ones, thirty-three percent of the subjects' VoI utility functions are more globally concave, seventeen percent of subjects' utility functions are indistinguishable across domains, and twenty-two percent of the subjects' utility functions differ in domain but we can not rank them by relative concavity. This provides a nonparametric classification of participants as efficiency preferring, equity preferring, or egoist, respectively. We re-estimate the reduced demand functions for both the allocation to Poor, $x_{\text{VoI}}(q,z)$, and Bad, $x_{\text{Risk}}(q,z)$, using this nonparametric classification. We still find that demand for $x_{\text{Risk}}(q,z)$ is uncorrelated with the assignment to social preference catagory.

In our third analysis of the data, we provide a structural estimation of a two-parameter subjective expected utility model at the individual level. One parameter reflects an individual's subjective prior for the Low state, which under the appropriate restriction reduces the model to an expected utility model, and the other measures the curvature of the power utility function. We find the majority of participants are optimistic by over-weighting the Good/Rich state. The estimated curvatures are generally higher for VoI than for Risk problems. This is mainly driven by the larger magnitudes of the right tail of VoI-selves' curvature estimates. Though highly noisy in terms of standard errors, the difference between individuals' estimated curvatures displays significant clusters of the three social preference types. We find seventeen percent of subjects' utility functions are significantly more concave for Risk than for VoI tasks, exhibiting equity preferring while thirteen percent subjects significantly show a taste for equity.

The rest of the paper is structured as follows. In Section 2.2, we discuss the related literature on measuring individual social preferences with induced budget experiments. In Section 2.3, we introduce the experimental design and procedures. Section 2.4 presents the standard statistical data analysis and estimates the reduced demand curves. In Section 2.5, we present our revealed preference analysis. Section 2.6 reports our structural estimation results. Section 2.7 concludes.

2.2 Literature review

Recent work has recognized the importance of estimating the distributional preferences of those who set policies and of those who follow them. For example, Fisman et al. (2015) and Li, Dow and Kariv (2017) conduct induced budget dictator game experiments respectively with medical and law students at prestigious universities. They presume these students are representative of future policy designers in health and other areas. In the domain of those who are governed by policies, Saez and Stantcheva (2016) and Kuziemko et al. (2015) elicit redistribution preferences from random samples of U.S. taxpayers. Then they explore how these preferences influence optimal taxation policy. Alternatively, Fisman, Jakiela and Kariv (2017) conduct an induced budget dictator game experiment with a random sample of American voters. They relate these preferences to voting behaviour and political party affiliation.

Most of these studies use a variation of the dictator game to elicit a social preference measure. More broadly, there are three prominent elicitation methods, each imposing an alternative DM perspective. These three perspectives are (1) as a dictator in front of a VoI, (2) as a disinterested social planner behind a VoI and (3) as a society member behind a VoI.[5]

A standard Dictator game has two players in which one decides how a fixed amount of money will be divided between herself and the other. Individuals typically do not claim the total amount for themselves and a large literature has identified multiple motivations for the finding. Some studies modify the standard version of the dictator game by changing the set of possible choices to include allocations which vary in both total aggregate wealth and wealth inequality. For instance, Charness and Rabin (2002) and Engelmann and Strobel (2004) limit the dictator's choice sets to a small number of wealth profiles, each corresponding to an alternative theory of social preferences. They establish that individual choices are driven by efficiency concerns as much as equity ones and then propose quasi-maximin preference models, loosely speaking weighted averages of efficiency and the maximin payoff, to explain their aggregate data. Cox (2004) provides a stronger demonstration of efficiency motives with a dictator game in which the two players are endowed with the same level of currency and the dictator can increase the wealth of the other player with a price of one-third; almost two-thirds of the dictators sent positive amounts of money. These studies effectively demonstrate

[5] Kariv and Zame (2014) propose an interesting theoretical model in which well behaved preference orderings corresponding to our Risk setting and a Dictator setting over the wealth profiles of our VoI setting are projections from a preference ordering in our VoI setting. Under assumptions of sufficient DM rationality and number of Risk and Dictator choices, one can recover the DM's VoI preferences.

that aggregate choices reflect consideration of efficiency in addition to that of inequity, but they do not provide the opportunity to explore the structure and heterogeneity of individual preferences.

A different strand of the literature examines the rationality and structure of individual dictator preferences through induced budget experiments where participants choose from multiple linear wealth profile possibilities frontiers. Andreoni and Miller (2002) is the first in this strand with a core treatment consisting of eight downward sloping budget sets with prices varying from one-third to three. Over ninety percent of their participants do not violate GARP, as compared to approximately twenty-two percent from a simulation of individuals who simply make choices according to a uniform distribution on a budget set. Moreover, they find three clusters of individual choice patterns: claim all available wealth, choose according to Leontief preferences (which is consistent with inequity aversion) and maximize aggregate wealth. Fisman, Kariv and Markovits (2007) introduce a graphical interface allowing for a rapid collection of decisions. In their design, participants complete fifty budget set dictator problems with prices mainly ranging from one-half to two. They find participants' choices are largely consistent with GARP, reflect similar behavioural types as found in Andreoni and Miller, and this is reflected in the various clustering of estimated utility functions in the family of constant elasticity of substitution forms. We should note that both of these two studies differ from ours in that each participant is paid for two randomly selected decisions, once using their dictator choice and once when they are the recipient of a dictator.

A limited number of studies have elicited social preferences using both the perspective of a disinterested social planner behind the VoI and financial consequences.[6] In treatments relevant to our study, Traub et al. (2005); Traub, Seidl and Schmidt (2009) present tasks in which participants rank twelve alternative income distributions for a five-person economy behind the VoI both as a disinterested social planner and as a society member. Both studies conclude that on average the disinterested social planner perspective yields more equity preferring choices. However at the individual level, there is heterogeneity in this comparative static of moving from the disinterested social planner to a society member; about half the participants become more equity preferring, one-quarter more efficiency preferring and the remaining quarter's preferences do not change.

[6] There are a number of studies using hypothetical scenarios and questionnaires adopting this perspective. For example, see Michelbach et al. (2003); Bernasconi (2002); Amiel and Cowell (2000); Amiel, Cowell and Gaertner (2009).

Hong, Ding and Yao (2015) reports on the only, as far as we are aware, induced budget experiment for a disinterested social planner behind the VoI. In their experiment participants are randomly matched into trios for each of twenty decisions with linear budget sets for a two-person society. The prices range from one to ten and the largest expenditure level is over four times the lowest level. A single decision is selected, then one member of each trio is selected to be the social planner and the other two members are randomly allocated the rich and poor positions; the social planner is given a fixed payment invariant to the budget set. Hong, Ding and Yao find the GARP consistency of participants' choices are in line with the previous dictator studies. They then estimate a CES-utility function in which an elasticity of substitution less than -1 is interpreted as efficiency preferring and a value greater than -1 as equity preferring. They report about sixty percent of their participants' estimated elasticities reflect efficiency preferring.

There is a large experimental literature which elicits individual social preferences from the perspective of a society member behind a VoI. We discuss the two most relevant studies which address the conflation of the risk attitude and the social preference behind the VoI. Frignani and Ponti (2012) presents an experiment in which a participant participates in twenty-four decision problems in which they select one two-element wealth profile from four possibilities. Their between-participant design has three treatments: a dictator in front of a VoI, a society member behind a VoI, and a lottery treatment where participants only see their own payoff-relevant information. Their analysis starts from the assumption that individuals maximize a mean-variance expected utility function and the variance term coefficient reflects the strength of inequality aversion in the dictator treatment, a mix of risk and inequality aversion in their behind the VoI treatment and risk aversion in their lottery treatment. They find the distribution of the coefficients for the last two treatments are similar and conclude that choices in the VoI treatment are largely determined by risk attitude. They also provide evidence that the distribution of the estimated variance coefficients has a lower mean than those of the other treatments. These results would appear to stand in contrast to ours. However, the between-subject design does not allow for the intrapersonal comparison of an individual's Risk and VoI choices which is key to identifying their social preference. Indeed we also find there is no treatment effect in the aggregate, but that is because the proportions of equity and efficiency preferring individuals are relatively balanced in our sample. Hence our within-subject design, and our much weaker assumptions on the form of utility, generates a more informative assessment of individual preferences.

Schildberg-Hörisch (2010) conducts an experiment in which a participant completes a pair of consumer problems like we pose, choosing a contingent claim portfolio and a wealth profile behind a VoI from the same budget set. However, she only considers a single budget set, which has a price of two. A participant receives three payments; the result of the Risk problem, then once as the decision-maker for the VoI and then again as the "other" society member for a different participants' VoI decision. She observes a difference between the average choices in the Risk and VoI problems, participants allocate more to the poor in the VoI than they do to the Low state in the Risk problem. These results are consistent with our findings, but with the lack of budget set variation there is no ability to evaluate individual-level preferences and the validity of the result is limited to the single price and expenditure level.[7]

Finally, the Risk problem component of our instrument is closely related and contributes to the literature on eliciting risk preferences using linear budget sets experiments. Seminal papers by Choi et al. (2007*a*,*b*) introduce an instrument by which individuals select contingent claim portfolios from a random sequence of fifty budget sets. These budgets sets have prices mostly ranging between one-half and two (states did not have the High-Low convention we adopted). Their study consists of treatments in which the states are equally likely and in which one state was twice as likely. While they find individual preferences are strongly consistent with GARP they are heterogeneous. However, many individuals' choices are not consistent with expected utility; for example not fully insuring when the price is one or selecting portfolios that violate first order stochastic dominance. To accommodate such observations they cleverly use a generalized version of the Gul (1991) disappointment aversion model. Choi et al. (2014) apply this instrument to a representative sample of the Dutch population to demonstrate its external efficacy and find that the level of consistency with GARP is lower than that found with U.S. college students. They also find this consistency is positively correlated with household wealth and negatively correlated with age.

2.3 Experimental design

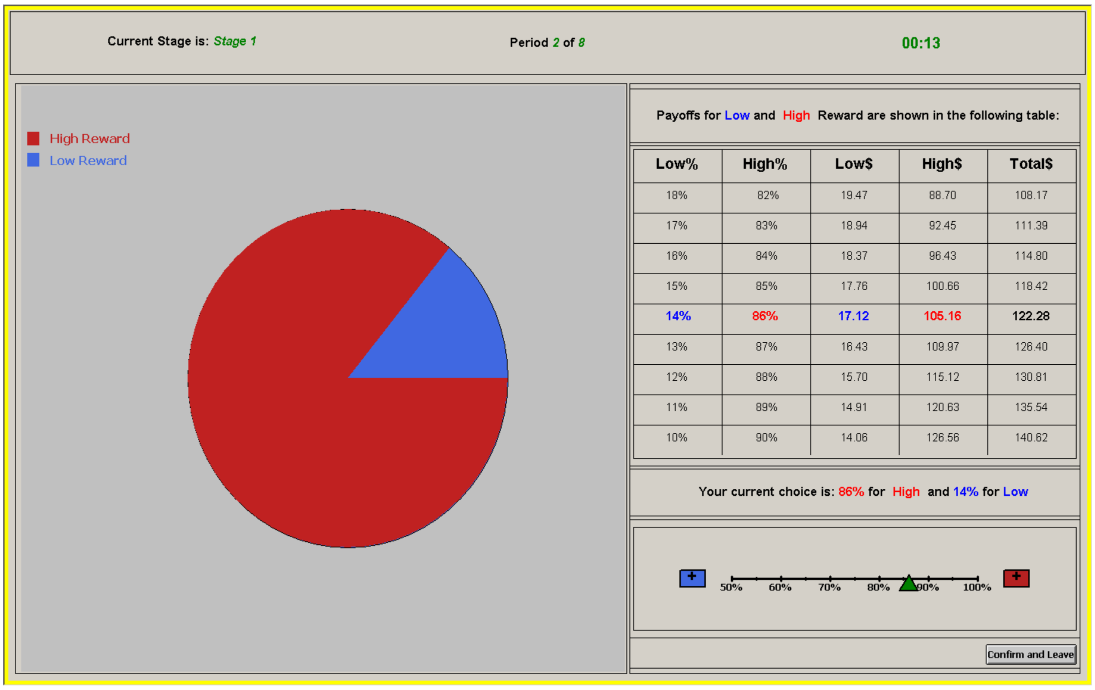
We conducted our experiment in the Financial and Experimental Economics Laboratory (FEEL) at Xiamen University. A total of 92 subjects were recruited via ORSEE (Greiner (2004)) and came from a broad range of majors. There were 9 sessions and each one lasted around 100 minutes. The average payment was 55 CNY, not including a 10

[7] Note this study also has a treatment in which participants make a pair of Dictator and VoI decisions. She finds the average amount given to the poor is less in Dictator problem than in VoI problem. Becker, Häger and Heufer (2013) report similar results based on many more rounds of choices.

CNY show-up fee. At the time of the experiment, the exchange rate was approximately 1 USD = 6.2 CNY.

A session consists of two stages: forty rounds each of a VoI and Risk problem. Each VoI problem has a matched Risk problem with the same budget set. We construct the forty budget sets by taking the Cartesian product of the set of prices $q \in \{1,2,3,4,5\}$ and the set of expenditures $z \in \{50,80,110,140,170,200,230,260\}$. We index all budget sets by (q^j, z^j) with $j = 1,\ldots,40$, and (x_i^j, y_i^j) denotes a chosen allocation from budget set j for choice problem i with $i \in \{\text{Risk, VoI}\}$.

We develop a new interface to implement our choice problems. Rather than choosing a point on a budget line, a participant selects a point using a dynamic pie chart and an accompanying table which she controls with a slider. Figure 2.2 shows the interface for a typical choice round. The pie indicates both the monetary amounts and shares allocated to High (red) and Low (blue) rewards. The table on the right hand shows the current and nearest divisions of the pie (Low% and High%), the amount of Low and High reward (Low$ and High$), and the current size of the pie (Total$). A participant adjusts allocations by moving the green triangle along the slider, which has a precision of 1%, to the desired position. When she adjusts the slider, the High/Low reward shares and the size of the pie adjust accordingly. Once the participant determines her most favored allocation for the current problem, she can click the “Confirm and Leave” button and start a new round.



Payoffs for Low and High Reward are shown in the following table:

Low%	High%	Low$	High$	Total$
18%	82%	19.47	88.70	108.17
17%	83%	18.94	92.45	111.39
16%	84%	18.37	96.43	114.80
15%	85%	17.76	100.66	118.42
14%	86%	17.12	105.16	122.28
13%	87%	16.43	109.97	126.40
12%	88%	15.70	115.12	130.81
11%	89%	14.91	120.63	135.54
10%	90%	14.06	126.56	140.62

Your current choice is: 86% for High and 14% for Low

Figure 2.2: A typical decision screen interface

To control for potential framing effects we use two frames to describe the tasks. In one frame the slider's range is 0% to 50%, and in the other frame it is 50% to 100%, indicating proportions allocated to Low or High reward respectively. We set a random initial slider position each time when a participant makes a choice. We also control for order effects by setting half of the sessions as in the sequence VoI-Risk, and the other half as Risk-VoI.

A session with the High reward framing and the Risk-VoI order proceeds through the following steps. First, we collect participants' informed consents. Second, a monitor reads the stage one instructions in Appendix 2.A.1 aloud while participants silently read along from a hard copy. After reading the instructions, we answer any questions and then participants have to successfully complete a simple quiz to ensure their comprehension. Third, participants complete their stage one decisions and then are given a short break. Fourth, after the break, we distribute the stage two instructions as in Appendix 2.A.2 and repeat the process. Notice that in this VoI stage instructions, we explain to the participant she will be randomly and anonymously matched with another participant in the same session. Before determining payoffs, we ask subjects to complete a short computerized survey.

We pay participants for only one of their eighty decisions, and we randomly select this decision round after the completion of the two stages.[8] If we select a Risk decision round, a participant experiences an individual fair coin toss which determines whether she receives the High or Low reward. If a VoI decision round is chosen, the computer first displays decisions made by both the participant and her counterpart and then randomly selects one of the two decisions to implement. After that, we proceed to determine who of the paired participants receives High/Low reward as if we have selected a Risk decision round.

2.4 Reduced form analysis

In this section, we evaluate whether participants choose different allocations for Risk and VoI problems. In particular, we assess whether aggregate and individual level data are consistent with one of the three types of social preferences: egoist, equity preferring, or efficiency preferring. Figure 2.3 provides an array of scatter plots for three subjects' choices and corresponding budget sets. Each of these subjects corresponds to one of the preference types. Subject 48, whose choices are shown in the top row of the array, is

[8] To avoid wealth effects in experiments, a common practice is to pay participants based on their decisions in a random round (Cox, Sadiraj and Schmidt, 2015; Azrieli, Chambers and Healy, forthcoming).

an egoist. She almost allocates the same amount of money in all forty paired problems. Subject 4's choices, in the middle row of the array, exhibit a taste for equity. In Risk problems, she allocates a vast majority of her expenditure to the Good state but always chooses nearly equal wealth distributions in VoI problems. Subject 24's choices, in the bottom row of the array, exhibit a taste for efficiency. In Risk problems, she always insures heavily even when the price is high, but then she selects lopsided wealth profiles in VoI problems. In general, most subjects' choices show a clear inclination towards one of the three categories. However, within the same category decision patterns are highly heterogeneous.[9]

We proceed by providing statistical evaluations of the treatment effects at multiple levels. In these evaluations, the dependent variable is the paired difference between amounts allocated to the Poor individual and to the Bad state, $x_{\text{VoI}} - x_{\text{Risk}}$, in the VoI and Risk problems respectively.

2.4.1 Aggregate and individual treatment effects

Before aggregating the data, we find neither framing nor order effects in participants' decisions.[10] At the fully aggregated level, we find no significant difference in allocations of Risk and VoI problems; just as Frignani and Ponti (2012). Table 2.1 reports means and standard errors of monetary and proportional amounts allocated to the Bad and Poor state. Across forty rounds of paired problems, participants on average allocate 30.38 to the Bad state and 30.26 to the Poor member; the difference is statistically insignificant according to a pairwise t-test and a Wilcoxon signed-rank test. Similarly, the average proportion chosen for Low reward account is 27.34% and 27.73% for the Risk and VoI treatments respectively; this difference is statistically significant.[11] However, in our judgment this is not economically significant. Since proportional allocations cannot fully capture the influence of varying budget sets, we focus on the monetary amounts going forward.

These null results at the aggregate level mask important heterogeneity among participants. Figure 2.4 reports the mean difference of $x_{\text{VoI}} - x_{\text{Risk}}$ and the corresponding 95% confidence interval for each participant. We plot these means and confidence intervals by stacking from the lowest mean to the highest and then capping the stack

[9] The scatter plots for all 92 participants' choices are available from authors on request.

[10] This result is confirmed by both the t-tests and Wilcoxon rank-sum tests on amounts allocated to the Bad state and the Poor individual. With respect to the framing effects, the p-value is 0.20 (0.22) for the t-test and is 0.09 (0.14) for the Wilcoxon rank-sum test on Risk (VoI) choices. For the order effects, the p-value is 0.17 (0.13) for the t-test and is 0.07 (0.32) for the Wilcoxon rank-sum test on Risk (VoI) choices.

[11] The p-value is 0.08 for the t-test and 0.02 for the Wilcoxon signed-rank test.

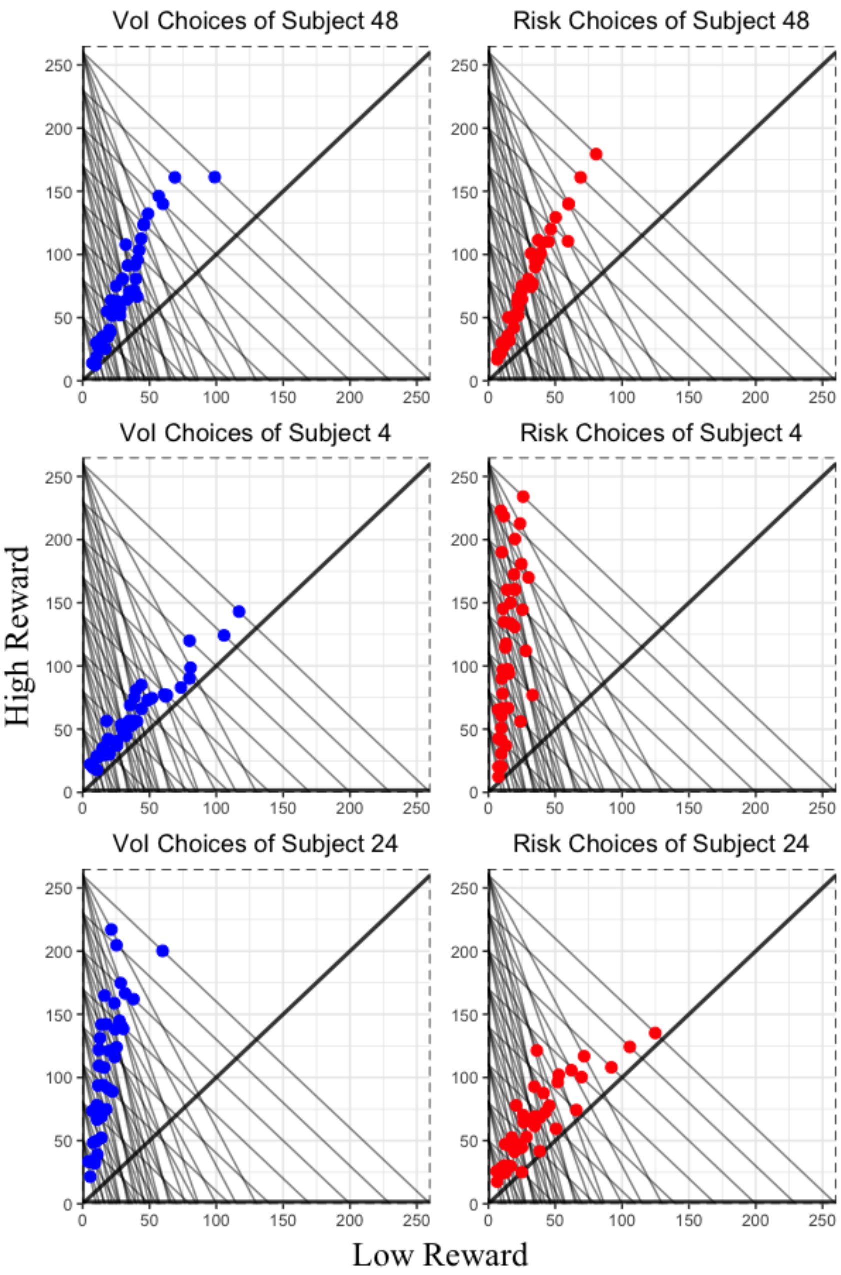


Figure 2.3: Illustrative examples for three types of choices: egoist (Subject 48), equity preferring (Subject 4) and efficiency preferring (Subject 24)

with the mean and confidence interval for the aggregate data. This figure allows us to identify three clusters of individuals. First, there are 51 participants who we can not reject as an egoist as their confidence intervals straddle zero. This cluster includes both those who are truly egoist and those for whom this hypothesis test result is a type two error; their unconditional choices have too much variation to precisely measure their non-zero mean. For now, we proceed to label this cluster as egoist. Second, there are

Variable	Monetary Allocation		Proportional Allocation	
	x_{VoI}	x_{Risk}	$x_{\text{VoI}}\%$	$x_{\text{Risk}}\%$
Mean	30.26	30.38	27.73	27.34
SD	26.09	25.75	15.45	14.69

Table 2.1: Summary statistics of choices in the VoI and Risk treatments averaged across budgets and individuals

24 participants whose confidence intervals lie above zero. This is evidence for equity-preferring, and we label this cluster Equity. Third, there is a cluster of 17 participants whose confidence intervals lie below zero. These participants' choices provide evidence for efficiency-preferring and we label this group Efficiency.[12]

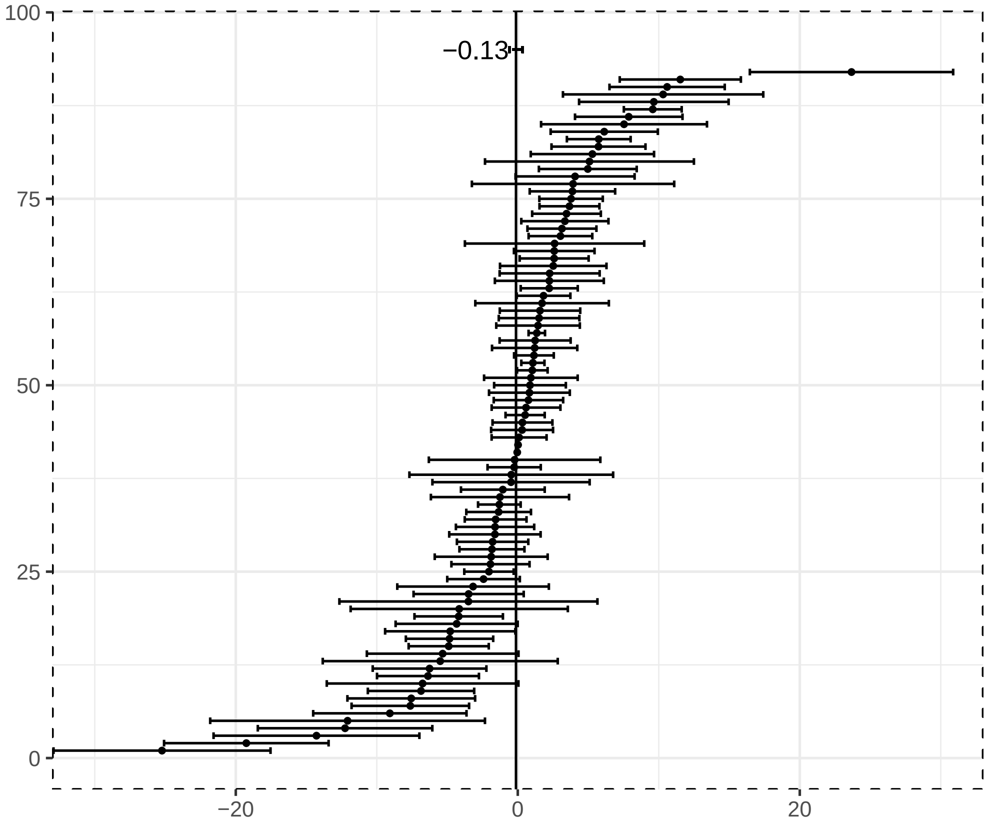


Figure 2.4: The mean and 95% confidence intervals of the paired difference between Poor and Bad state allocations, $x_{\text{VoI}}(q,z) - x_{\text{Risk}}(q,z)$, for each participant. These confidence intervals are calculated based on paired t-tests for each individual. We stack them from the lowest mean to the highest. The confidence interval for the aggregate data is then stacked at the top.

[12] In Appendix 2.B, we provide group classifications based on eight alternative criteria.

2.4.2 Reduced form demand estimation

We estimate reduced form demand functions for the Bad/Poor allocation for each of the three groupings — Equity, Egoist, and Efficiency. Table 2.2 reports estimated coefficients and robust standard errors clustered at the individual level. We first estimate a simple model where the dependent variable is the demand of the monetary allocation to the Low state, x_i, and the independent variables are the constant, price, and expenditure. We interact each of the terms with a dummy variable for the VoI tasks, D_{VoI}, to create three new factors. The original coefficients are estimates for the Bad state demand function and the interaction terms' coefficients are the estimated deviations for the Poor allocation demand function. We then run a second set of models in which we add a dummy variable $D_{p=1}$ for the price when it equals one. For such choice problems, the price of insurance is actuarial fair in the Risk treatment and there are one-for-one wealth transfers between rich and poor, ergo aggregate wealth is constant, in the VoI treatment. We speculate that behavior takes an abrupt change at this price. This speculation is justified as the intercept terms are greatly reduced, price coefficients more reasonable, and the new coefficients are highly significant for all three groupings. We proceed using the results of this second set of models.

There is evidence that individuals' risk preferences are uncorrelated with their efficiency-equity preferences. Inspection of the estimated coefficients for the Risk treatment shows that they are very close in value for the three sub-groupings of social preferences. We conduct an F-test to evaluate the joint null hypothesis that the three groups' reduced demand functions in the Risk treatment are the same. We fail to reject this null with an F-statistic of 11.79, which is asymptotically χ^2 distributed with eight degrees of freedom and has a p-value of 0.16. This implies that there is no correlation between the social preference categorization and the attitude towards risk.

The coefficients of the Risk demand functions are sensible. Increasing the price of the Bad-state contingent claim by one dollar leads a strong re-balancing of portfolios towards more wealth in the Good-state. Expenditure coefficient estimates also suggest that for every increase of 30, our expenditure step size, only 6 will be used to purchase Bad-state contingent claims and 24 will be added to the Good-state component of the portfolio.

There are notable differences between the estimated VoI demand curves of the Efficiency and Equity groups. First, and almost inevitable given the construction of the sub-samples, the price intercept is greater for the Equity group's VoI demand relative to that of the Efficiency group. Surprisingly when we fix expenditure, the Efficiency group's VoI demand curve is less price sensitive than the Equity group's one. However,

	Dependent variable: x_{VoI} and x_{Risk}					
	Equity	Egoist	Efficiency	Equity	Egoist	Efficiency
Constant	28.32**	29.79**	29.45**	11.34**	11.90**	14.90**
	(1.77)	(1.40)	(2.33)	(1.29)	(1.35)	(1.20)
Price	-9.57**	-10.06**	-10.24**	-5.32**	-5.59**	-6.60**
	(0.76)	(0.58)	(0.98)	(0.52)	(0.45)	(0.59)
Expenditure	0.20**	0.20**	0.21**	0.20**	0.20**	0.21**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Constant*D_{VoI}	5.13**	0.08	-9.06**	5.13**	0.08	-9.06**
	(1.21)	(0.97)	(2.28)	(1.21)	(0.97)	(2.28)
Price*D_{VoI}	-1.93**	0.11	3.71**	-1.93**	0.11	3.71**
	(0.42)	(0.41)	(0.91)	(0.42)	(0.41)	(0.91)
Expenditure*D_{VoI}	0.04**	-0.004	-0.07**	0.04**	-0.004	-0.07**
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
$D_{p=1}$				21.22**	22.37**	18.20**
				(2.03)	(2.08)	(2.49)
Observations	1,920	4,080	1,360	1,920	4,080	1,360
R^2	0.72	0.62	0.58	0.72	0.62	0.58
Adjusted R^2	0.72	0.62	0.58	0.72	0.62	0.58
Note:	*p<0.05; **p<0.01					

Table 2.2: Estimated reduced form demand curves for Low/Poor wealth allocations by group classification. OLS regressions with robust standard errors clustered at the individual level. The classification is based on two-sided *t*-test with 95% confidence level.

when expenditure increases the Equity group's generosity increases at a greater rate than the Efficiency group's. With an expenditure step of 30, the Equity group uses about 7.2 to create a transfer to the Poor, while the Efficiency group uses only 4.2.

The statistical analyses show that there are roughly three types of social preference categories. However, even within the same type, especially the Egoist type, there is a large variation in individual choices. For some subjects, their confidence intervals are quite narrow and we may safely categorize them as egoists. However, many subjects' mean choice differences are consistent with one type but are erroneously classified as egoist due to their wide confidence interval. Some subjects may respond more strongly to expenditure variations or to price variations leading to these wider confidence intervals. To better capture these latent structures, and avoid the myriad of other questionable assumptions underlying these analyses, we utilize nonparametric revealed preference techniques in Section 2.5 below.

2.5 Nonparametric analysis

We readdress how effectively our instrument elicits efficiency-equity preferences using nonparametric revealed preference techniques. We first test whether participants' choices, made by both Risk- and VoI-selves, are rational in terms of maximizing a monotonic and concave utility function, and if so, whether that function is homothetic. We also check the power of our experimental design and compare it with the power of other induced budget experiments. We then evaluate the relative concavity of each participant's VoI and Risk utility functions and classify participants into different social preference types. We re-estimate the reduced form demand functions according to this classification to further validate our result on the independence of risk and social preference.

2.5.1 Testing rationality and homotheticity

We test consistency of each participant's choices with GARP to address the question of rationality.[13]

For price taking DMs and linear budget sets, Afriat (1967) and Varian (1982) show that satisfying GARP is both a necessary and sufficient condition for the existence of a continuous, monotonic, and concave utility function that is maximized by the observed choices. This condition imposes no parametric structure on choice data and is easily testable based on implicit revealed preference relationships. It is a commonly employed test for induced budget experiments.[14]

It is common to find GARP violations in choice data. In response, researchers have proposed various ways to measures the extent choices violate GARP. Afriat (1967) propose the Critical Cost Efficiency Index (CCEI), the minimum scalar multiplier one needs to apply to the set of expenditure levels to resolve all GARP violations of a DM. The CCEI is bounded between zero and one. The higher the CCEI of a DM's choices,

[13] In our experiment, when a DM chooses (x^j, y^j) from a budget with price q^j and expenditure z^j she directly reveals that she prefers (x^j, y^j) over all other (x^k, y^k) for which $q^j x^k + y^k \leq z^j$. If the inequality is strict (i.e., $q^j x^k + y^k < z^j$) we say that (x^j, y^j) is strictly directly revealed preferred over (x^k, y^k). Furthermore, if she chooses (x^k, y^k) from the budget with q^k and z^k, she indirectly reveals that she also prefers (x^j, y^j) to all bundles affordable under q^k and z^k. This is how the indirect revealed preference relation is constructed through the transitive closure of the direct revealed preference relation. GARP states that if a choice (x^j, y^j) is (directly or indirectly) revealed preferred to a choice (x^k, y^k), then (x^k, y^k) must not be strictly directly revealed preferred to (x^j, y^j). Varian (1982) shows that GARP is equivalent to the existence of a continuous, monotonic, and concave utility function u which rationalizes the choices, that is, $u(x^j, y^j) \geq u(x, y)$ whenever (x^j, y^j) is revealed preferred over (x, y).

[14] Some notable examples are Cox (1997); Sippel (1997); Harbaugh, Krause and Berry (2001); Andreoni and Miller (2002). A recent theoretical justification for the validity of using experiments to test utility maximization via GARP was provided by van Bruggen and Heufer (2017).

the more rational she is.[15] The CCEI can be interpreted as a measure for wasted income: A DM with a CCEI of e could have obtained the same level of utility by spending only a fraction of e of what she actually spent to obtain that level of utility.

We also test choices for consistency with maximizing an increasing homothetic utility function.[16] We do this for two reasons. First, while we believe making allocation choices via a dynamic pie chart is more intuitive than via a graphed budget line, it potentially makes two decision heuristics focal: constant High/Low pie shares and constant budget shares.[17] When a DM uses either heuristic, her choices necessarily satisfy GARP. Second, it is an important validation check when one estimates homothetic utility functions using data from induced budget experiments. For example, Andreoni and Miller (2002) and Fisman, Kariv and Markovits (2007) estimate CES utility functions and Choi et al. (2007*a*) estimate CRRA utility functions.

Following Heufer (2013), we evaluate participants' choices for consistency with the Pairwise Homothetic Axiom of Revealed Preference (PHARP).[18] Heufer (2013) shows that this condition is equivalent to the constrained maximization of a homothetic utility function in two commodity choice settings. We also calculate Heufer and Hjertstrand's (2017) Homothetic Efficiency Index (HEI) can be interpreted in analogy to the CCEI as a measure of wasted income on resolving violations of PHARP.

While efficiency indices have appealing interpretations, there is no natural scale to measure irrationality. When a DM makes "errors," her efficiency level will depend on the number of budget sets and the variations in both price and expenditure. Bronars (1987) suggests an exercise that simultaneously provides a measure of power for a sequence of linear budget sets and benchmarks for alternative rationality thresholds for an efficiency index. One simulates a large number of DM choice sequences — we will use 10,000 — in which each choice is a random draw from the uniform distribution over the budget line. Then one constructs an empirical density of the resulting efficiency indices. For a given rationality threshold, the proportion of simulated sequences exceeding it is a measure of the corresponding power. Also, the densities of indices from the simulated data serve as benchmarks for those of actual DMs. We report two power calculations:

[15] We say that (x^j, y^j) is directly revealed preferred over (x^k, y^k) *at efficiency level e* if $q^j x^k + y^k \leq e z^j$, where $e \in [0,1]$. GARP at efficiency level e can be defined accordingly. The CCEI of an individual is the greatest efficiency level e at which that individual's choices satisfy GARP.

[16] A utility function is homothetic if it is a positive monotonic transformation of a utility function that is homogeneous of degree one.

[17] Choices made according to a constant High/Low pie share heuristic lie on a single ray through the origin. Choices made using a constant budget shares heuristic lie on a family of rays through the origin, each ray corresponding to a particular price ratio.

[18] PHARP states that for any distinct pair of choices, indexed by j and k, $(q^j x^k + y^k)(q^k x^j + y^j) \geq z^j z^k$ holds.

one using uniform random draws from budget lines, and another using uniform draws restricted to the part of the budget above the 45° line. The latter restriction mimics the restriction imposed on the participants in our experiment.

2.5.2 Rationality results and comparisons with other studies

We now present the results for rationality checks and the power of our experimental design, and compare these with various induced budget experiments.

Rationality and homotheticity of participants' choices

By problem type, we check every possible pair of a participant's choices for GARP violations and count them.[19] The test is binary: a choice pair either satisfies GARP or not. Overall, the number of GARP violations is small, with a median of 3 and 4 respectively for the Risk and VoI problems. Over 30% of participants commit zero or only one violation of GARP. There are 11 participants who fully satisfy GARP in the VoI problems and 12 in the Risk problems; 4 satisfy GARP in both problem types. With respect to the axioms regarding homotheticity, there are 3 participants in both the VoI and the Risk problems who satisfy PHARP; only one of them satisfies HARP in both problem sequences.[20]

Next, by problem type, we calculate the CCEI and HEI for each participant and report summary statistics for the resulting distribution of values in the first two rows of Table 2.3. The mean CCEIs are 0.96 and 0.95 for the Risk and VoI problems respectively, and the median CCEIs are both 0.98. Also, 73%, 87%, and 99% of the participants' Risks (VoI) CCEIs exceed the respective thresholds of 0.95, 0.90, and 0.80.[21] While the majority of participants commit at least one GARP violation, there is compelling evidence almost all are highly rational. Most participants' violations are not severe because small adjustments in expenditures brings full compliance with GARP. In contrast, the HEI results suggest that many participants do not make choices consistent with homothetic preferences. The average HEI is 0.85 in both problems, and only about one-third of the HEIs, 32% and 36% for Risk and VoI problem respectively, exceed the threshold of 0.90.

[19] For each treatment, there are 780 pairs of such choices.

[20] We do not report violation counts for PHARP as it only considers pairwise violations; a count of violations is therefore not comparable to a count of GARP violations.

[21] There is no consensus in the literature regarding the thresholds for efficiency indices. For example, Andreoni and Miller (2002) uses a CCEI threshold of 0.95 for their dictator game with eight budgets, while Fisman, Kariv and Markovits (2007) omit participants with a CCEI below 0.8 from further analysis in their dictator game with fifty budgets.

The power of the collection of linear budget sets

We report on the distributions of the two efficiency indices generated by Bronars's suggested simulation — adjusted for the budget sets of our instrument. These distributions provide more context to our claims of high rationality and an evaluation of the power of our linear budget sets. Figure 2.5 compares the distribution of CCEI for our 92 participants and for 10,000 hypothetical participants who choose randomly from the part of the budget above the 45° line. The consistency levels of the participants are overwhelmingly higher than those of random choice. Figure 2.6 similarly compares the distribution of HEI between real and simulated participants. While the HEI values of the participants are considerably lower than their CCEI values, the figure shows that the participants still have much higher homothetic efficiency than the random choices.

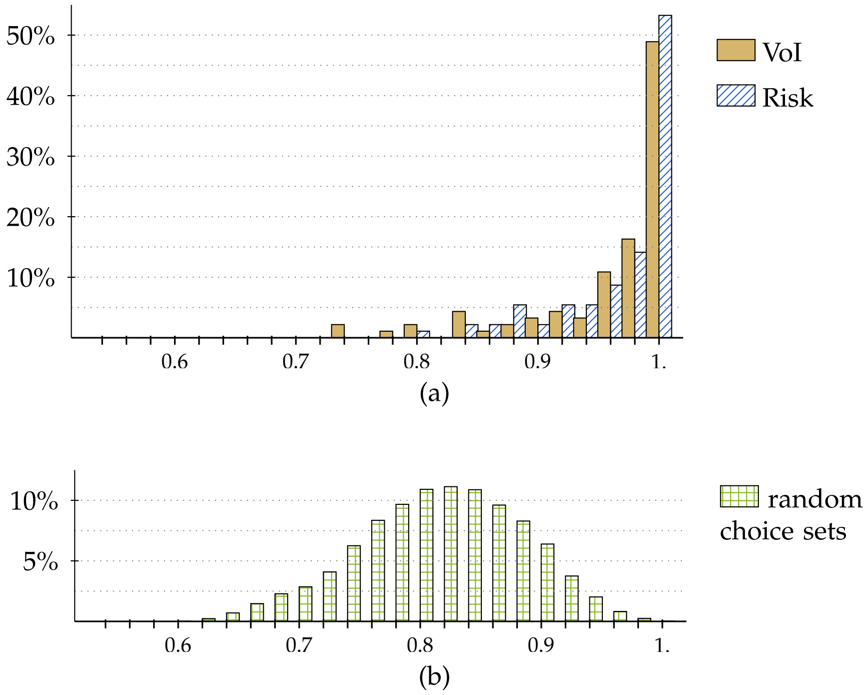
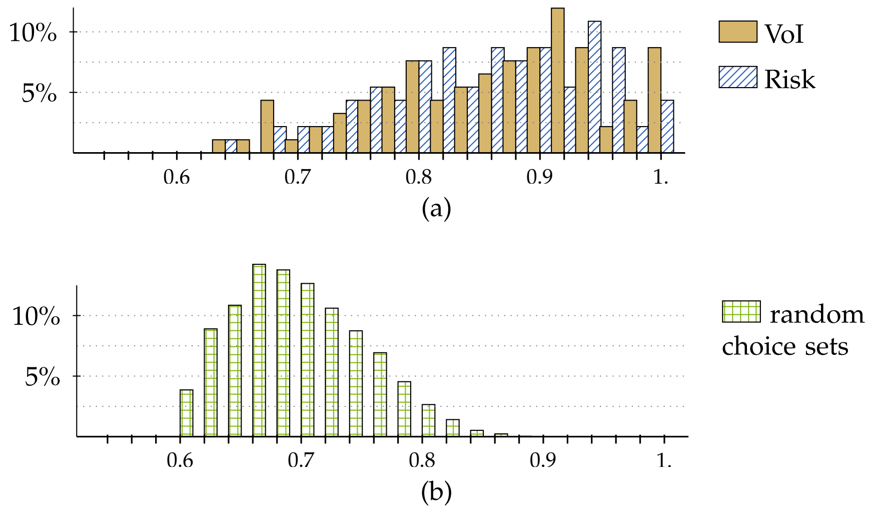


Figure 2.5: Histogram of the CCEI (a) for 92 subjects and (b) for 10,000 random choices from uniform distributions over the portion of the budget set above the 45° line.

The third row of Table 2.3 reports summary statistics for the distributions of CCEIs and HEIs generated by simulated choices. These values are strikingly lower than those for the Risk and VoI problems. The simulated HEI rarely exceed any of our benchmark thresholds. In the fourth row we report the same summary statistics for simulations of random draws from a uniform distribution over the portion of the budget line above the 45° line. There is little evidence of this censoring leading to critically low test



(a)

(b)

Figure 2.6: Histogram of the HEI (a) for 92 subjects and (b) for 10,000 random choices from uniform distributions over the portion of the budget set above the 45° line.

power, except perhaps for the CCEI at the 80% benchmark. We confidently rule out the possibility of random behaviour generating the observed high consistency levels.

			CCEI					HEI				
Study	Experiment	Decision maker	Mean	Median	≥ 0.95	≥ 0.90	≥ 0.80	Mean	Median	≥ 0.95	≥ 0.90	≥ 0.80
Current	Risk	Subjects	0.96	0.98	73%	87%	99%	0.85	0.86	11%	32%	72%
	VoI	Subjects	0.95	0.98	73%	84%	95%	0.85	0.87	13%	36%	70%
		Random	0.75	0.75	0%	3%	30%	0.60	0.59	0%	0%	0%
		Random above 45°	0.80	0.81	1%	7%	53%	0.68	0.68	0%	0%	2%
Choice under risk	CFGK	Subjects	0.94	0.97	68%	81%	89%	0.83	0.86	19%	38%	68%
		Subjects censored	0.96	0.98	79%	87%	94%	0.85	0.87	19%	40%	72%
		Random	0.65	0.66	0%	0%	6%	0.52	0.52	0%	0%	0%
		Random above 45°	0.83	0.84	1%	16%	71%	0.70	0.71	0%	0%	9%
	CKMS	Subjects	0.88	0.93	45%	58%	76%	0.78	0.80	22%	33%	50%
		Subjects censored	0.95	0.98	67%	81%	96%	0.86	0.87	28%	43%	70%
		Random	0.72	0.74	1%	6%	30%	0.58	0.58	0%	0%	1%
		Random above 45°	0.89	0.90	19%	48%	88%	0.75	0.76	0%	2%	30%
Dictator giving	FKM	Subjects	0.96	0.99	71%	87%	97%	0.88	0.92	41%	55%	82%
		Random	0.60	0.61	0%	0%	4%	0.47	0.47	0%	0%	0%
	AM	Subjects	0.99	1.00	98%	99%	99%	0.98	1.00	83%	94%	99%
		Random	0.87	0.90	37%	50%	74%	0.76	0.77	4%	12%	41%

Table 2.3: Comparison of efficiency indices and test power results for various comparable experiments: FKM (Fisman, Kariv and Markovits, 2007), AM (Andreoni and Miller, 2002), CFGK (Choi et al., 2007*a*) and CKMS (Choi et al., 2014).

Test power and efficiency levels compared to other experiments

We compare our efficiency indices and corresponding power assessments with those of other similar budget experiments, namely the decision under risk studies of Choi et al. (2007*a*; CFGK) and Choi et al. (2014; CKMS) as well as the dictator studies of Andreoni and Miller (2002; AM) and Fisman, Kariv and Markovits (2007; FKM). These experiments, including ours, differ in the subject pool, the number of budgets, and the price and expenditure variations; all of which may affect economic indices and test powers. In particular, our experiment restricts participants' choices to be above the 45° line, which may deprive chances of violating GARP and PHARP and thus result in higher indices. To better evaluate this concern, we also additionally impose similar restrictions to the data of CFGK and CKMS. Note that CFGK considers two different priors on the Good state of the world; probabilities ⅔ and ½. We only include data on the latter in our analysis. Further note that in CFGK and CKMS either good could be the cheaper one, while in our framework, good *x* is never cheaper than *y*. Accordingly we first transform CFGK's and CKMS's data by labeling the cheaper good as *y*. We then censor choices below the 45° line by replacing them with a choice on the 45° line, thereby "correcting" choices that violate first order stochastic dominance (FOSD).[22]

Overall our experiment's power and participant's rationality are in line with these other experiments. With respect to the choice under risk studies, inspection of the middle rows of Table 2.3 reveals our participants' CCEIs are slightly higher for the mean, median and all threshold levels when compared to the original CFGK and CKMS data, and slightly lower when compared to the censored data. As we ex-ante believed our design would nudge participants to be more homothetic, is it somewhat surprising that our participants do not universally exhibit higher homothetic efficiency; the differences are relatively minor. The power of all of these designs is generally very high.

The results for the two dictator games are quite different due to the large difference in the number of budgets; FKM asked participants to make 50 choices, while AM's participants only made 8 choices. Accordingly, for the CCEI, the results for FKM are similar to our results for the VoI part. With respect to the HEI, our participants do considerably worse. Part of the reason is that many participants in dictator games keep all or almost all of the money to themselves which reduces the chance of violating

[22] For dictator games, this restriction does not make sense, as ungenerous dictators have no reason to allocate higher amounts to the other recipient even if it is relatively cheap to do so. Therefore, we do not impose this restriction on FKM and AM. Also when we report power calculations for linear budget dictator game studies we only report on random selection with the full budget line as the support.

PHARP. Our participants have lower efficiency compared with AM, which can be explained by the fact that the power of our experiment is substantially greater.

2.5.3 The relative convexity of Risk-self versus VoI-self preferences

We return to the idea of classifying individual participants' social preferences into the categories of equity preferring, efficiency preferring, and egoist. We refine these classifications based on a nonparametric approach proposed by Heufer (2014). This approach allows the evaluation of the relative global convexity of two families of indifference curves when each is used to generate optimal choices against the same budget sets. We use this approach to evaluate the relative convexity of the family of indifference curves — i.e. concavity of the utility function — for each participants' Risk- and VoI-selves. Note that the more concave a DM's Risk-self utility function is relative to that of her VoI-self, the smaller the difference between her allocation to the Poor individual in a VoI problem and her allocation to the Bad state in a Risk problem when facing the same budget set.

Application of this approach requires an additional rationality condition on a set of linear budget set choices: a DM's choices must be consistent with *second order stochastic dominance* (SOSD). A lottery (wealth profile) A has SOSD over a lottery (wealth profile) B if and only if every expected (average) utility maximizer prefers A over B (Hadar and Russell, 1969). A DM's choices are consistent with SOSD if B is never revealed preferred over A whenever A has second order stochastic dominance over B.[23]

We check each of our participant's choices, by problem type, for consistency with SOSD-GARP and compute the SOSD-CCEI as defined by Heufer (2014), which are the SOSD analogies to GARP and the CCEI. We find that 11 (9) subjects in the Lottery (VoI) treatment satisfy SOSD-GARP compared to 12 (11) subjects who satisfy GARP. The SOSD-CCEI values are exactly the same as the CCEI values for all but three subjects who satisfy GARP but not SOSD in one of problem types. This is not surprising as subjects were required to make choices above the forty-five degree line, making it difficult to violate stochastic dominance without also violating GARP. As in Heufer (2014) we use efficiency-adjusted revealed preference relations in order to meet the condition.

The key concept for our approach is a revealed binary relation we call Partially Revealed More Convex (PRMC), where convex refers to the shape of the indifference

[23] Note that consistency with SOSD does not require consistency with expected utility.

curves.[24] We say a DM is Risk-PRMC at an observed Risk choice, denoted A, if there exists at least one allocation, denoted B, from the DM's set of VoI choices such that

(i) the DM's expected wealth level from the wealth profile B is higher than her expected wealth of A;
(ii) A is revealed preferred to B by the Risk-self;
(iii) B is revealed preferred to A by the VoI-self; and,
(iv) at least one of those two revealed preference relations is strict.

VoI-PRMC is analogously defined.

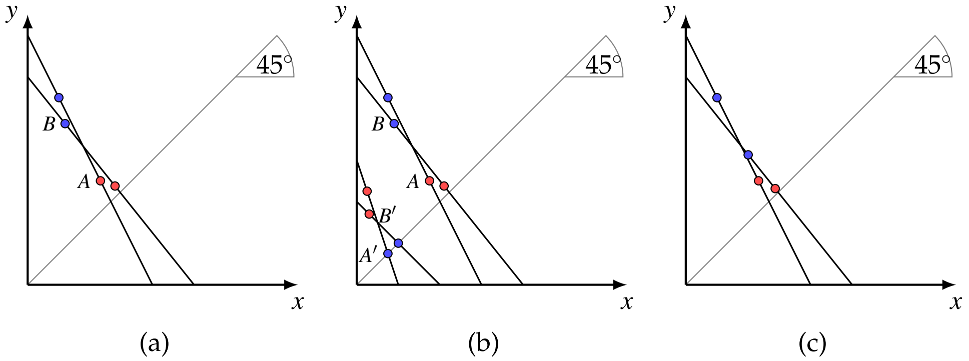


Figure 2.7: Examples of PRMC relationships. Choices made in Risk and VoI problems are in red and blue, respectively. In (a), the DM is Risk-PRMC at the observed Risk choice bundle A. In (b), the DM is Risk-PRMC at Risk choice bundle A and VoI-PRMC at the VoI choice bundle A'. In (c), the DM's choices show neither Risk-PRMC nor VoI-PRMC.

Figure 2.7a illustrates a Risk-PRMC relationship through a DM's choices in Risk and VoI problems. The DM's Risk-self prefers A over B even though B has a higher expected value supposedly because B is too risky for the DM's Risk-self. In other words, compared to the Risk-self, the VoI-self is "more risk-seeking" since she is willing to accept a "more risky lottery" that delivers a higher aggregate wealth (i.e., is more efficient) and this extra "risk-seeking" component reflects a taste for efficiency. Therefore, passing the Risk-PRMC test at A establishes that the indifference curve of DM's Risk-self is more convex than that of her VoI-self at A. The test for VoI-PRMC is analogously defined. If a DM passes the VoI-PRMC test at an observed choice from a VoI problem, her VoI-self is willing to take a more equal but less efficient allocation than her Risk-self, and therefore the indifference curve for VoI-self is more convex at this particular choice.

[24] The original approach in Heufer (2014) was developed to compare the risk aversion of two individuals. PRMC corresponds to Heufer's partially revealed more risk averse (PMRA) relation.

For each participant we test eighty PRMC relations, forty each for Risk and VoI choices. Each test result is an observation-specific binary indicator, which jointly provides us a global picture of the relative convexity of Risk and VoI-self's preferences. We impose the most strict requirement on these 80 test results and refine our social preference classifications into four categories:

(i) unorderable: both Risk-PRMC and VoI-PRMC hold at least once;
(ii) efficiency-preferring: Risk-PRMC holds at least once while VoI-PRMC never holds;
(iii) equity-preferring: VoI-PRMC holds at least once while Risk-PRMC never holds;
(iv) egoist: both Risk-PRMC and VoI-PRMC never hold.

In three out of the four cases, we reach a definitive conclusion about a DM's global social preference implied by a straightforward reinterpretation of Theorem 3 in Heufer (2014). Case (i) is the inconclusive one in which we can conclude a DM's Risk-self and VoI-self utility functions are distinct but there is no global relative rank of their concavities. Figure 2.7b provides an example of such un-orderable Risk- and VoI-self preferences. The DM is Risk-PRMC at the Risk choice A, revealing a taste for efficiency (locally); while she is VoI-PRMC at the VoI choice bundle A', revealing a taste for equity (locally). The un-ordable preferences imply that either (1) people have different notions of the trade-offs in Risk and VoI problems, or (2) people's preferences are not stable over the entire expenditure range. In case (ii) the DM's Risk-self choices can be represented by a more concave utility function than the VoI-self choices but not vice versa. Case (iii) is the reverse. Case (iv) implies that the same utility function can represent both selves. For instance, in Figure 2.7c, we cannot find a PRMC relationship based on the DM's choices. When all her choices follow such a pattern, the underlying indifference curves for her Risk- and VoI-self are globally similar.

Table 2.4's first row summarizes the classification results. Out of our 92 participants, 76 participants (83%) demonstrate a non-trivial social preference and the other 16 participants (17%) are classified as egoist. Among participants exhibiting social preferences, 26 clearly exhibit a taste for efficiency; 30 are equity-preferring; the other 20 participants have distinct Risk- and VoI-self utility functions but we can't establish a global concavity ranking. Compared to our earlier type clusters based on t-tests, participants in the group of efficiency-preferring and equity-preferring largely overlap. However, the revealed preference approach allowed us to refine many of the those who we labeled as egoist into a pair of clusters: those who are egoist and those having a social preference which is not globally consistent with respect to trade-offs between efficiency and equity. Thus, we provide some resolution to the issue of identifying those individuals for whom the difference between their Bad and Poor quantities demanded is zero from those individuals for whom the difference is not zero but the sign differs by budget sets — and is a possible source of the instances of wide confidence intervals around zero in Figure 2.4.

Expenditure slack	Different			Similar
	Equity preferring	Efficiency preferring	Un-orderable	Egoist
full slack	30 (32.6%)	26 (28.3%)	20 (21.7%)	16 (17.4%)
.9999	30 (32.6%)	26 (28.3%)	20 (21.7%)	16 (17.4%)
.975	16 (17.4%)	22 (23.9%)	7 (7.6%)	47 (51.1%)
.95	13 (14.1%)	13 (14.1%)	1 (1.1%)	65 (70.7%)

Table 2.4: Nonparametric catagorization of social preference types

Notice our nonparametric classification results are based on participants' pre-adjusted revealed preference relations. Before examining the PRMC relation, we slack each participant's budget by SOSD-CCEI to resolve SOSD-GARP violations. What if a participant is Risk-PRMC (or any other non-egoist type) based on the full slack, but no longer when there is a slight disturbance on the adjustment? We test the robustness of our classification results with different expenditure slacks when correcting for SOSD-GARP violations. When there is a small noise in the expenditure adjustment, the classification results are replicated in the second row of Table 2.4. As the expenditure slack gets smaller, the difference between Risk- and VoI-self preferences are required to be stronger to pass the PRMC tests. If a DM is persistently classified in one of the different preference groups, we are more certain about her preference type in terms of economic significance. The third and fourth row of Table 2.4 show that when the slack index decreases, more participants are classified into the egoist preference group. With a slack index of 0.95 (which corresponds to the standard significance level of 0.05 in economics), we have around 30% participants who demonstrate different preferences for Risk and VoI problems.[25]

[25] In unreported analyses, we adjust the budgets on an alternative efficiency index, the Varian Efficiency Vector (Varian, 1993; Heufer and Hjertstrand, 2017). Compared to the CCEI efficiency index, an upper bound for the adjustment for all budgets to resolve GARP (or SOSD-GARP) violations, VEV corrects the revealed preference relation whenever it is necessary, thus keeping more information implied in people's choices. We follow the same procedure and re-classify our subjects into four preference types. Compared to the classification results in Table 2.4, more subjects (93.4%) demonstrate distinctive preferences for Risk and VoI problems. All three non-trivial social preference types (equity-preferring, efficiency-preferring, and un-ordable) also show greater tolerance for the expenditure slack. With expenditure slack of 0.95, 20 (21.7%) participants are equity-preferring; 15 (16.3%) are efficiency-preferring; and 16 (17.4%) participants' Risk- and VoI-self preferences are un-ordable.

2.5.4 Social preference classification stability

Our nonparametric social preference classifications overcome some of the issues associated with our previous statistical approach. In particular, it allows refinement of the egoist classification into those whose preferences are globally similar and those whose preferences are not comparable. Further, it relaxes potentially invalid assumptions underlying the statistical tests used to generate classifications. To evaluate the impact of these factors, we show how individual participants' classifications differ under the two approaches. Then we examine how the nonparametric approach impacts estimates of type specific reduced form demand functions.

We first examine how participants' classifications vary under the two approaches. Table 2.5 lists the number and fraction of subjects who are classified into different social preference types under both the statistical and nonparametric approach. The rows are the classification results under the statistical t-test, and the columns are those under the nonparametric test. We first note participants with preferences classified as un-orderable originally come from all three preference types. These are pairwise not significantly different from each other under the proportion test. This counters our expectation that those classified as un-orderable would have previously been classified as egoist. We secondly note there is little switching of participants between Equity Preferring and Efficiency under the two classification approaches. However, subjects who are categorized as Egoist under t-test are now almost equally distributed among three social preference clusters. This result confirms that (1) the three social preference types elicited by our experimental instrument are robust; (2) our nonparametric method provides a more refined classification result.

	Nonparametric categorization				
	Equity Prefer	Efficiency Prefer	Egoist	Un-orderable	Sum
Equity Prefer	15 (88.2%)	0 (0.0%)	2 (11.8%)	7	24
Efficiency Prefer	1 (7.7%)	11 (84.6%)	1 (7.7%)	4	17
Egoist	14 (33.3%)	15 (35.7%)	13 (31.0%)	9	51
Sum	30	26	16	20	92

Table 2.5: The cross table for the categorization under the statistical and nonparametric approach. The rows are the classification results based on two-sided t-test, with 95% confidence interval. The fractions in the parentheses are calculated by excluding subjects who are classified as un-orderable.

We conclude this section by reporting in Table 2.6 the re-estimation of reduced form demand functions for each social preference type under the nonparametric classification.

Note that the interaction terms with the treatment dummy are less significant compared to the estimation based on t-test. This is due to that two-thirds of the subjects who were previously egoist are now assigned to the equity and efficiency group, reducing the precision of our treatment effect estimates. However, a joint F-test for the reduced demand functions for the Risk treatment across three clusters (excluding the un-orderable subjects) shows that they are not significantly different. The F-statistic is 13.08, and the p value is 0.11 with a degree of freedom of 8. At the same time, a joint F-test for the reduced demand functions for VoI treatment across three clusters (excluding the un-orderable subjects) shows that they are significantly different. The F-statistic is 40.71, and the p value is 0.00 with a degree of freedom of 6. Thus, our initial evidence that social preference types form distinct clusters but risk preferences are independent of these types persists in this alternative nonparametric approach.

	Dependent variable: x_{VoI} and x_{Risk}			
	Equity	Egoist	Efficiency	Un-orderable
Constant	11.51**	14.93**	11.99**	11.80**
	(1.00)	(2.02)	(2.08)	(1.89)
Price	−5.24**	−7.51**	−5.34**	−5.43**
	(0.46)	(0.57)	(0.70)	(0.60)
Expenditure	0.19**	0.23**	0.20**	0.18**
	(0.01)	(0.01)	(0.01)	(0.01)
Constant*D_{VoI}	2.44*	−2.41	−3.49	1.44
	(0.99)	(1.90)	(1.97)	(2.18)
Price*D_{VoI}	−0.76	1.36*	1.15	−0.34
	(0.47)	(0.64)	(0.77)	(0.91)
Expenditure*D_{VoI}	0.02*	−0.02	−0.03*	0.002
	(0.01)	(0.01)	(0.01)	(0.01)
$D_{p=1}$	18.84**	24.61**	20.21**	23.75**
	(2.05)	(3.12)	(2.61)	(3.38)
Observations	2,400	1,280	2,080	1,600
R^2	0.65	0.68	0.60	0.59
Adjusted R^2	0.65	0.68	0.60	0.59
Note:				*p<0.05; **p<0.01

Table 2.6: Estimated reduced form demand curves for Low/Poor wealth allocations by group classification. The classification is based on the non-parametric approach. OLS regressions with robust standard errors clustered at the individual level.

2.6 Structural equation estimation and analysis

Estimates of structural utility models are useful in counter-factual policy analysis when the assumed utility model accurately approximates DMs' preferences. In this section, we measure the efficiency-equity trade-offs through a parametric approach. We estimate a two-parameter subjective expected utility model for each participant. One parameter reflects the relative subjective prior of the Bad/Poor state. The other reflects the curvature of the utility function. We find the estimates of subjective priors are consistent with overconfidence in the Rich/Good state of the world. We also show that individual differences between estimated Risk and VoI utility curvatures form clusters of social preferences similar to those found in our reduced form and nonparametric analyses. Finally, we find a similar range of curvature estimates, for both Risk and VoI tasks, as Choi et al. (2014) but greater estimates than Choi et al. (2007*b*).

2.6.1 Subjective expected utility specification and estimation approach

We assume a DM's choices for a decision round with relative price q and expenditure z are governed by the solution of the constrained maximization problem

$$\begin{aligned} \underset{(x,y)}{\arg\max} \quad & \alpha u(x) + u(y) \\ \text{subject to} \quad & qx + y = z \\ & y \geq x \geq 0. \end{aligned} \tag{2.1}$$

The parameter α is the ratio of her subjective priors of the Low and High state of the world, ${}^{\Pr(\text{Low})}/_{\Pr(\text{High})}$.[26] When $\alpha = 1$, the DM has an objective prior in which two states are equally likely, and consequently, is an expected utility (EU) maximizer. When $\alpha \neq 1$, the DM's prior differs from the objective one and she is a subjective expected utility (SEU) maximizer. For $\alpha > 1$ she believes the low state is more likely and vice versa when $\alpha < 1$. Note, a DM's subjective prior of the Bad/Poor state is $\alpha/(1+\alpha)$.[27]

[26] Notice for the VoI task, we are following the arguments of Harsanyi (1953, 1976) that a social welfare function closely approximates to an average expected utility function.

[27] An alternative interpretation for α is a relative weight the DM attaches to a state with a less favorable outcome while attaining the objective probability over the two states. When $\alpha > 1$, she over-weights the Low state due to disappointment aversion (or loss aversion if assuming equal allocations as reference points) as described in Gul (1991), and vice versa when $\alpha < 1$.

We assume the utility function for the same DM is the commonly employed power utility function (see Wakker (2008)),

$$u(x) = \frac{x^{1-\rho_i}}{1-\rho_i},$$

where $i \in \{\text{Risk}, \text{VoI}\}$. The parameter ρ_i defines the concavity of the DM's problem specific utility function. For the Risk choice problem ρ_{Risk} is a constant coefficient of relative risk aversion. In terms of empirical execution, the power function is not well-defined for boundary allocations $(x_i, y_i) = (0, z)$. However, these choices might be realized when there is an unobserved noise component in the DM's choices. When evaluating these corner choices we replace 0 with a small number ωz, where $\omega = 0.001$.[28]

Solving the maximization problem (2.1) yields the following equation characterizing the optimal commodity bundle choice,

$$\ln(x_i^*/y_i^*) = f(q, \omega; \alpha, \rho_i) = \begin{cases} \ln(\omega) & \text{if} \quad \ln(\alpha) - \rho_i \ln(\omega) \leq \ln(q), \\ -\frac{1}{\rho_i}\left[\ln(q) - \ln(\alpha)\right] & \text{if} \quad \ln(\alpha) < \ln(q) < \ln(\alpha) - \rho_i \ln(\omega), \\ 0 & \text{if} \quad \ln(q) \leq \ln(\alpha). \end{cases}$$

The first case is the corner solution $(x_i^*, y_i^*) = (\omega z, z)$; the second case is an interior solution and the last case is for a corner solution in which $x_i^* = y_i^*$. We use a dummy variable formulation to indicate whether a decision round j is either a Risk or VoI problem. With respect to the utility curvature parameter ρ we define,

$$\rho_{\text{VoI}} = \rho_{\text{Risk}} + D_{\{j=\text{VoI}\}} \cdot \rho_{\text{diff}}.$$

The parameter ρ_{diff} is the difference of the curvature between two utility functions: one recovered from the DM's VoI-self and the other from her Risk-self. With positive ρ_{diff}, the DM is globally more equity preferring and a negative ρ_{diff} implies efficiency-preferring.

We obtain estimates of the DM's utility parameters by minimizing the square sum of the distances between $\ln(x/y)$ and $\ln(x^*/y^*)$, where the former vector is composed of her choices in eighty decision rounds and the latter from the corresponding optimal choices. Specifically, we solve the problem

[28] All following estimation results are robust to reasonable alternative choices of ω.

$$\underset{\alpha,\rho_{\text{Risk}},\rho_{\text{diff}}}{\arg\min} \sum_{i\in\{\text{Risk, VoI}\}} \sum_{j=1}^{40} \left[\ln\left(\frac{x_i^j}{y_i^j}\right) - \ln\left(\frac{x_i^{j*}}{y_i^{j*}}\right)\right]^2 =$$

$$\sum_{i\in\{\text{Risk, VoI}\}} \sum_{j=1}^{40} \left[\ln\left(\frac{x_i^j}{y_i^j}\right) - f\left(q^j, \omega; \alpha, \rho_{\text{Risk}}, \rho_{\text{diff}}\right)\right]^2.$$

We estimate belief and curvature parameters by the nonlinear least square (NLLS) method using the Levenberg-Marquardt (Moré, 1978) algorithm.[29]

2.6.2 Individual level parametric estimation

We estimate structural models for 64 (70%) participants and exclude 28 participants. First, we omit 6 participants whose choices are not sufficiently consistent with utility maximization as indicated by CCEI values of less than 0.8 for either Risk or VoI treatment. Second, we exclude subjects who fail to make choices varying enough to ensure convergence. For instance, 6 participants chose at least 15 boundary allocations in which $x = 0$ in both VoI and Risk problems; 2 participants chose too many boundary allocations in which $x = y$; and 14 participants either frequently distributed fixed proportions or amounts to one asset/wealth component, regardless of the budget set.

We report a summary of estimation results in Table 2.7.[30] By subject, we estimate both an expected utility, by restricting $\alpha = 1$, and a subjective expected utility specification. For the expected utility estimations, we report summary statistics of the individual estimates of ρ_{Risk} and ρ_{diff}, and the calculated values of ρ_{VoI}. For the subjective utility estimations we additionally report results on the individual estimates of α.[31][32]

[29] As with many optimization algorithms, this algorithm only guarantees the convergence upon local optima. Accordingly, we first run the OLS on the interior optimal condition to get reasonable starting parameters for the NLLS estimation. For subjects who choose few boundary allocations, the initial OLS estimation will be close to the region where the global optimum lies.

[30] The formulation of the dummy variable for a Risk/VoI problem allows us to jointly estimate the DM's utility parameters for Risk- and VoI-selves. The joint estimation procedure implicitly assumes two error terms associated with Risk and VoI problems respectively follow the same distribution, which is reasonable under our experimental design of matched pairs of forty decision rounds. In unreported results, We have estimated utility parameters for Risk and VoI problems separately, and the resulting percentiles are identical to those reported in Table 2.7.

[31] In unreported results, we also estimated another version of the subjective expected utility model in which a DM's α can vary between the Risk and VoI problems. For 62 out of 64 participants, we cannot reject the null hypothesis of the same prior parameter at a 5% significance level.

[32] We attempted to conduct the same procedure for the constant absolute risk aversion (CARA) utility function. We do not report these results as the numerical optimization algorithm did not find parameter

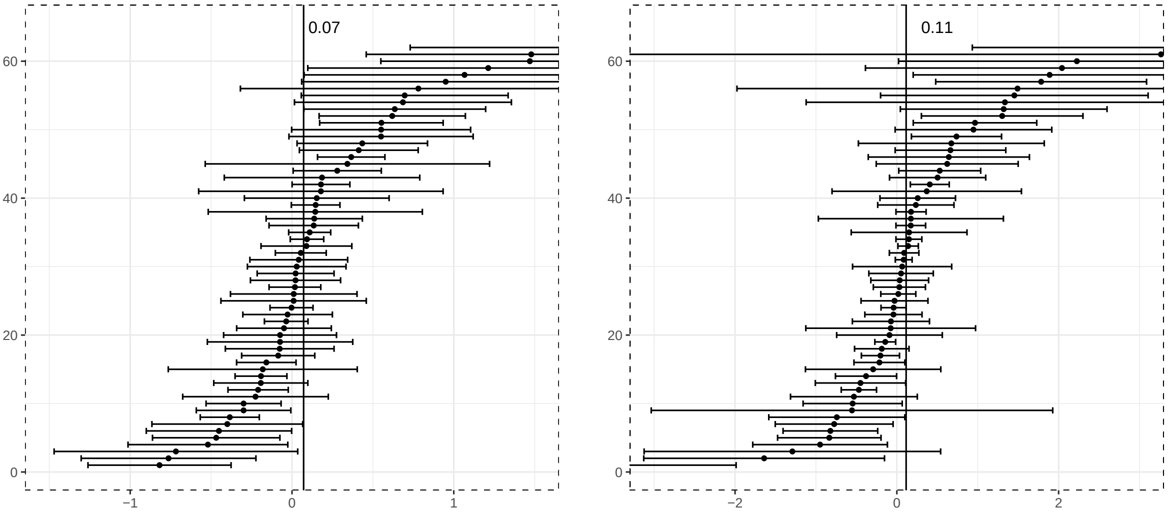
We note several observations from Table 2.7. First, for both EU and SEU, the mean of estimated utility curvatures is greater for Risk than VoI. The mean estimates of ρ_{Risk} are 1.16 and 2.77 for EU and SEU respectively, and the corresponding mean estimates of ρ_{VoI} are 1.61 and 3.81. Second, the variance of estimated curvatures is higher for the VoI than Risk choice problems - for both EU and SEU specifications. This is driven by the right tail of the distribution of ρ_{VoI} estimates. The median, the twenty-fifth and the fifth quantiles are virtually the same for ρ_{Risk} and ρ_{VoI} and for higher quantiles we find ρ_{VoI} exceeds ρ_{Risk}. Third, for the SEU specification, 87.5% (56 out of 64 participants) of the estimated α's are smaller than one, reflecting optimism for the higher payoff. Since the EU specification is nested within SEU, we use a log-likelihood ratio test to evaluate the hypothesis a participant is an EU-maximizer against the alternative of a SEU-maximizer. For 42 out of 64 subjects, we reject EU model at the 5% level of significance.

	Expected Utility ($\alpha = 1$)			Subjective Expected Utility ($\alpha > 0$)			
	ρ_{Risk}	ρ_{VoI}	ρ_{diff}	α	ρ_{Risk}	ρ_{VoI}	ρ_{diff}
Mean	1.16	1.61	0.45	0.53	2.77	3.81	1.04
Variance	0.34	4.46	3.55	0.25	8.69	40.25	27.42
p5	0.44	0.44	-0.51	0.01	0.34	0.47	-0.93
p25	0.73	0.72	-0.10	0.09	1.09	1.15	-0.21
p50	1.04	1.16	0.07	0.46	1.80	1.82	0.11
p75	1.49	1.67	0.46	0.76	3.17	4.20	0.69
p95	2.16	2.92	1.48	1.51	7.12	12.12	3.11

Table 2.7: Summary of NLLS estimation of individual level parameter estimates of ρ_i and α. The quantile a is denoted as pa.

For both utility specifications, we evaluate whether a participant has the same preferences in the Risk and VoI choice problems by testing whether their estimated ρ_{diff} is statistically different from zero. We present a visualization of the hypothesis tests results in the pair of stacked confidence interval plots presented in Figure 2.8. Despite the lack of precision in many estimates, as demonstrated by the numerous wide 95% confidence intervals, we find more evidence supporting the same heterogeneity of social preferences found in earlier reduced form and non-parametric analyses. For the EU specification, 26 out of 64 subjects' ρ_{diff} are significantly different from zero: sixteen larger and ten smaller than zero. For the SEU specification, 20 out of 64 subjects' ρ_{diff} are significantly different from zero: eleven larger and nine smaller than zero.

estimates for a large majority of subjects. The estimation results based on a sub-sample of convergence subjects are consistent with those with CARA utility functions and comparable with other experiments.



(a)

(b)

Figure 2.8: The estimated curvature differences ρ_{diff} based on (a) expected utility and (b) subjective expected utility model for each participant. We stack them from the lowest to the highest. Also notice that estimates from two subjects from each model are excluded because their values and confidence intervals are out of figures' ranges. The median ρ_{diff} for all 64 subjects is 0.07 in EU and 0.11 in SEU model, and both are stacked at the top.

2.6.3 Comparison with other experiments

Next we report and compare structural estimates from the data of two previously considered induced budget set experiments preferences: CFGK and CKMS.[33] This meta-analysis is designed to identify how structural estimates can vary across subject pools and choice interfaces. To do this we censor or filter data in different ways to make the variety of choice sets comparable across studies.

We start with "Full" data sets which are not censored, but for CFGK and CKMS we relabel the state-contingent assets so the cheaper one is the x good. Next, for CFGK and CKMS, we replace choices that violate first order stochastic dominance (FOSD) with the fully insured portfolio on the budget line. We also consider "Price" adjusted data sets, including ours, by selecting decision rounds for which relative prices range between one and three. Then we consider simultaneously FOSD censored and price filtered

[33] We do not conduct structural estimates for the formally considered linear budget dictator experiments, AM and FKM, because participants knew the realized states and therefore the (subjective) expected utility framework is not applicable.

"FOSD & Price" data. We estimate the EU and SEU specifications for each individual and report the quartiles of estimate in Table 2.8.[34]

First, we find the curvature estimates from our experiment are much larger than those from CFGK but more comparable with those from CKMS. A closer examination of the estimate of the subjective prior parameter reveals that our participants are the most optimistic, CKMS's follow, and those in CFGK approximately assign equal probabilities for both states. As CFGK and CKMS share the same experimental protocols, the distinctions of both curvature and weight estimates may largely be attributed to the difference in subject pool: the former was conducted among US college student and the latter among a representative sample of the Dutch population.

How sensitive are the estimations to choice censoring, smaller price variation, and both? The corrections for first order stochastic dominance result in minimal changes in the curvature estimates of CFGK and CKMS but do increase the estimates of the subjective prior ratio of CKMS. Filtering for smaller relative prices has a differential impact across studies. In our study, estimates of ρ increase while those of α increase, both slightly. In CFGK ρ estimates increase but are almost unaffected in CKMS. In sum, both types of manipulation on choices affect the parametric estimation results to some degree, but they exhibit no unambiguous direction across three studies and furthermore, they do not abruptly alter the baseline magnitude of the estimates of each study.

Overall we find larger estimates for the curvature of power utility functions and smaller estimates of the subjective prior ratio between Low/High states than those from CFGK, but these estimates are more comparable with CKMS. These distinctions are more likely to be driven by different subjects pools rather than different experiment protocols.

[34] Note the constrained utility maximization problems for FOSD censored data sets are the same as in problem 2.1. For the full and the price only adjusted data sets, we cannot impose the second constraint as in problem 2.1 that the money allocated to y is higher than to x account. Hence the optimal conditions for demands of the two commodities are slightly different. Details can be found in Appendix 2.C.

			Expected Utility ($\alpha = 1$)			Subjective Expected Utility ($\alpha > 0$)					
Study	Experiment	Sample	ρ			α			ρ		
			p25	p50	p75	p25	p50	p75	p25	p50	p75
Current	Risk	Full	0.73	1.04	1.49	0.09	0.46	0.76	1.09	1.80	3.17
	VoI		0.72	1.16	1.67				1.15	1.82	4.20
	Risk	Price	0.71	0.99	1.24	0.23	0.53	0.71	1.10	1.75	2.76
	VoI		0.68	1.00	1.44				1.07	1.80	2.88
Choice under risk	CFGK	Full	0.31	0.58	0.79	0.77	1.01	1.24	0.25	0.48	0.93
		FOSD	0.31	0.59	0.84	0.91	1.06	1.17	0.30	0.52	0.96
		Price	0.36	0.66	0.94	0.80	0.95	1.05	0.36	0.72	1.18
		FOSD & Price	0.36	0.70	0.99	0.91	1.01	1.11	0.36	0.75	1.01
	CKMS	Full	0.65	1.03	1.51	0.45	0.78	1.01	0.78	1.27	2.17
		FOSD	0.69	1.13	1.72	0.73	1.00	1.22	0.75	1.12	1.86
		Price	0.54	0.91	1.56	0.50	0.82	1.00	0.69	1.28	2.64
		FOSD & Price	0.62	1.11	2.05	0.85	1.02	1.20	0.57	1.06	2.08

Table 2.8: Meta-comparison of structural estimates of EU and SEU specifications: Risk, VoI, CFGK (Choi et al., 2007*a*) and CKMS (Choi et al., 2014).

2.7 Conclusion

In this paper, we construct an instrument to identify and decompose social preferences behind a VoI. In particular, we examine pairs of contingent claim portfolio and VoI income distribution problems with a common budget constraint. A fixed commodity bundle induces the same marginal distribution of a DM's wealth in both choice problems. This allows us to use a DM's chosen contingent claim portfolio as a benchmark to establish whether her wealth distribution choice reflects equity preferring, efficiency preferring, or egoist social preferences. Our instrument consists of a set of forty linear budget sets for paired choice problems.

We applied this instrument in an experimental lab setting. Aggregate analysis suggests that there is no difference between behavior in the Risk and the VoI choice problems. However, our within-subject design allows us to demonstrate this is a fallacy. The analysis of intrapersonal differences points to clusters of social preference types. We establish the robustness of this conclusion using three empirical methodologies: reduced form statistical analysis, revealed preference analysis, and structural estimations of individual subjective utility functions. We found consistent classifications of individuals into groups exhibiting equity preferring, efficiency preferring, and egoist preferences.

Our results are informative to the heterogeneity of social preferences, with respect equity and efficiency, to the population of our sub-sample. However, we caution extrapolating these conclusions to other populations. We believe the intuitive interface and the structure of the data provided by our instrument lends itself well to more applied contexts. We intend to use it to measure the social and risk preferences of policymakers and beneficiaries in the developing world where the impacts of efficiency-equity trade-offs are acute. We also intend to use the instrument to measure risk and social preferences in organizational contexts where such information is key to develop effective contract designs. Finally, we are developing extensions to the instrument incorporating income mobility.

Appendix 2.A Experimental instructions

2.A.1 Instructions for the first stage

Preliminary Remarks

You are participating in an experiment investigating individual decision-making. Contingent on your decisions in this experiment, you can earn money in excess of your show-up fee of 10 RMB. Please pay careful attention to the instruction as a considerable amount of money is at stake.

During the experiment, please turn off your cell phones, laptop computers and other communication equipment. Do not turn on any software on the desktop other than the experimental application. Please do not communicate with any of the other participants or look at their computer monitors. If you have questions anytime, please raise your hand and we will address it as soon as possible. If you do not obey these rules, we will ask you to leave without any payment.

Overview

In today's experiment, you are going to complete 80 decision periods. The 80 periods are broken into two stages of 40 periods each—Stage 1 and Stage 2. There is a break of 10 minutes between two stages. At the end of the 80th period, only one decision period will be randomly selected as payment period. Therefore, your decision in each period might determine your final payoff. In today's experiment, there is a 50% chance you will receive a High Reward and a 50% chance you will receive a Low Reward. Whether you receive the High or Low reward will be determined only at the end of the experiment. Your decisions can't influence these probabilities; however, you can influence the amount of money associated with these rewards.

In a Stage 1 period, you will simply be asked to divide a pie of money between the High and Low Rewards. This division only affects the potential amount of money you will receive. Every other participant will make their own respective divisions, and whether you receive a High or Low reward has no bearing on what reward they receive. At the end of experiment, if the randomly selected period is in Stage 1, each subject will toss their own individual coin to determine whether they receive the High or Low Reward. Note that this determination will only be made if a Stage 1 period is selected

All monetary amounts in the experiment are Yuan. At the end of the experiment, you will be paid privately. When we call your ID number, please come forward to the sign-in counter and receive your earnings in the experiment plus your show-up fee. We ensure you that your participation in the experiment, your name, ID number and any information about your payoffs are confidential.

How to make decisions?

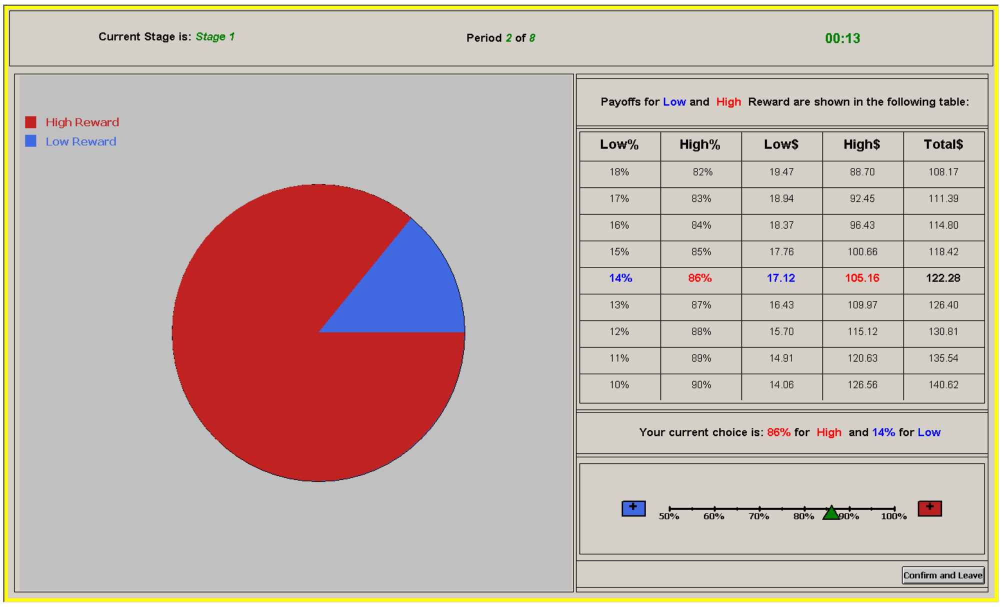
In each period in Stage 1, you will be asked to choose your most preferred division of a pie of money between the High (indicated by red) and Low (indicated by blue) Rewards by moving a slider. Your may receive the money from either High Reward or Low Reward as your final payment.

Your division is restricted in that High Reward must be assigned at least 50% of the pie and Low Reward no more than 50%. It is important to note that, the pie size (total reward) is not fixed and grows as you allocate a higher percentage of the pie to High Reward. In different periods, the pie size and growth rate vary. You should also note that even in the same period, the rate that pie size grows is not the same for different divisions. For example, the following hypothetical table shows how pie size change differently for the same 5% increase of share for two different divisions. For a (44%, 56%) division, a 5% increase of the High Reward share increases the pie size by $8.97, and for the (21%, 79%) division the same 5% increase of the High Reward share increase the total pie size by $16. To understand the pie growth rate you need to check the pie size at various divisions each period.

Low%	High%	Low$	High$	Total$
44%	56%	70.21	89.36	159.57
39%	61%	65.73	102.81	168.54
21%	79%	44.37	166.90	211.27
16%	84%	36.36	190.91	227.27

Table 2.9: Example of Varying Growth Rate for Same Proportional Increase

Figure 2.9 gives an example of what your computer screen may look like in one period. The top of the screen is a status bar showing current stage, current period and time elapsed in this screen. The left part is a pie indicating shares allocated to High (red) and Low (blue) Rewards. When you make your division choice, shares for High and Low in the pie will change accordingly, as will the pie size. The top of the right part is a 10 row table shows the current and the nearest divisions of the pie (Low% and High%), the amount of Low Reward and High Reward (Low$ and High$) and current pie size (Total$). The right bottom is the slider where you make decisions. The calibration of slider in this screen is 50% – 100%, which means points on the slider are proportions for High Reward. You can make your decision by either dragging the green triangle along the slider to your most desired position or clicking ”+” button at the two ends of the slider. Once you are satisfied with your division, click on the ”Confirm and Leave” button. After that, you will leave the current period and won’t be able to change your decision in this period. So please confirm your decision before clicking it.



Current Stage is: *Stage 1*

Period *2* of *8*

00:13

Payoffs for Low and High Reward are shown in the following table:

Low%	High%	Low$	High$	Total$
18%	82%	19.47	88.70	108.17
17%	83%	18.94	92.45	111.39
16%	84%	18.37	96.43	114.80
15%	85%	17.76	100.66	118.42
14%	86%	17.12	105.16	122.28
13%	87%	16.43	109.97	126.40
12%	88%	15.70	115.12	130.81
11%	89%	14.91	120.63	135.54
10%	90%	14.06	126.56	140.62

Your current choice is: 86% for High and 14% for Low

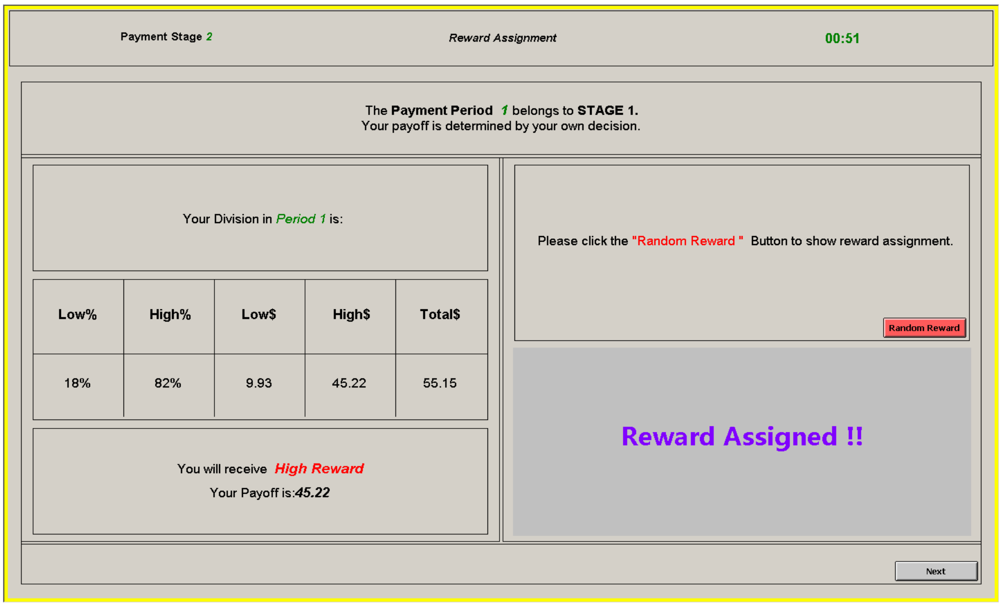
Figure 2.9: Decision Screen

How are earnings determined?

At the end of the experiment, the computer will randomly select one out of 80 periods for all subjects as Payment Period. Every period is equally likely to be chosen and the period selected is same for all subjects.

If the Payment Period belongs to Stage 1, the computer will randomly assign you either the High or Low Reward. Then your individual earnings will be determined according to the division you chose in the Payment Period.

Figure 2.10 shows the screen you will see if the Payment Period belongs to Stage 1. The top of the screen shows the Payment Period and which stage it belongs to. After you click ”Random Reward” button, the computer select the Reward Amount, displaying your Reward level and payoff. After that, a ”Next” button will appear and you can click it to leave this screen. In the example shown in Figure 2.10, the Payment Period is Period 1 and the reward is High, according to his division in Period 1 (18%, 82%), his payoff is $45.22.



Payment Stage 2

Reward Assignment

00:51

The **Payment Period** 1 belongs to **STAGE 1.**
Your payoff is determined by your own decision.

Your Division in *Period 1* is:

Low%	High%	Low$	High$	Total$
18%	82%	9.93	45.22	55.15

You will receive ***High Reward***
Your Payoff is:***45.22***

Please click the "Random Reward " Button to show reward assignment.

Reward Assigned !!

Figure 2.10: Payment Screen in Stage 1

2.A.2 Instructions for the second stage

Review

In today's experiment, you are going to complete 80 decision periods, which are broken into two stages—Stage 1 and Stage 2. At the end of the 80th period, only one decision period will be randomly selected as payment period. Your payment is a reward and there is a 50% chance you will receive a High Reward and a 50% chance you will receive a Low Reward. Whether you receive the High or Low reward will be determined only at the end of the experiment. Your decisions can't influence these probabilities; however, you can influence the amount of money associated with these rewards.

Overview

In a Stage 2 period, your task is still to divide a pie of money between High and Low Rewards. This stage is different because you will be matched with another participant in the experiment (called your counterpart), but we will not reveal who. Since you and your counterpart both make divisions of High and Low Reward, we only select one of two divisions to determine payoffs for the pair. There is a 50% chance your division will be carried out (and your counterparts' not) and 50% of chance your counterpart's division will be carried out (and yours not). This stage also differs in that one of you will receive the High Reward and the other the Low Reward. So there are two possibilities, each equally likely: (1) you receive the High Reward and you counterpart the Low

Reward, and (2) you receive the Low Reward and your counterpart the High Reward. Again, note that the division selection and reward determination will only be made at the end of the experiment if a Stage 2 period is selected.

How to make decisions?

In each period in Stage 2, you will be asked to choose your most preferred division of a pie of money between the High (indicated by red) and Low (indicated by blue) Rewards by moving a slider. The allocation you choose may determine the payoff of yourself and your counterpart. If one of the pair receives High Reward, the other one will get Low Reward surely.

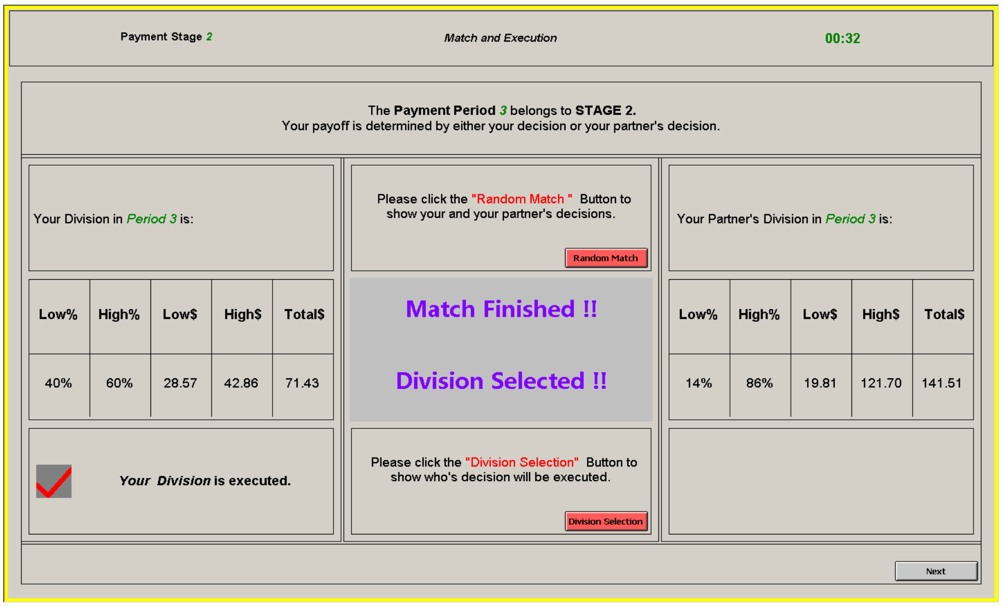
The interface you face is similar to that of Stage 1. The pie size (total reward) is not fixed and grows as you allocate a higher percentage of the pie to High Reward. In different periods, the pie size and growth rate also vary. Even in the same period, the rate that pie size grows is not the same for different divisions. You need to check the pie size at various divisions to understand the pie growth rate you each period.

How are earnings determined?

At the end of the experiment, the computer will randomly select one out of 80 periods for all subjects as Payment Period. Every period is equally likely to be chosen and the period selected is same for all subjects.

If the Payment Period belongs to Stage 2, the payment process will contain three parts. You will firstly be randomly paired with another participant. Then the computer will randomly select one of your two division choices to use to determine the money associated with the High and Low Reward amounts. Next, the computer will randomly assign the High and the Low Reward amounts to you and your counterpart.

Figure 2.11 shows the screen you will see if the Payment Period belongs to Stage 2. The top of the screen shows the Payment Period and which stage it belongs to. After you click ”Random Match” button, the computer will end the random match process and show both of your chosen divisions. After you click the ”Division Selection” button, the computer will end the random process to select one of the divisions, and a red check will appear under the selected table. In the example shown in Figure 3, your division will be carried out. It means one of you will receive $28.57 and the other one $42.86. After you click ”Next” button on the right bottom of the screen, you will see a similar screen for random reward. When you click ”Random Reward” button, you can see who was assigned the High Reward and who the Low Reward, as well as the corresponding payoffs.



The **Payment Period** 3 belongs to **STAGE 2.**
Your payoff is determined by either your decision or your partner's decision.

Your Division in Period 3 is:

Low%	High%	Low$	High$	Total$
40%	60%	28.57	42.86	71.43

✓ ***Your Division* is executed.**

Please click the "Random Match " Button to show your and your partner's decisions.

Match Finished !!

Division Selected !!

Please click the "Division Selection" Button to show who's decision will be executed.

Your Partner's Division in Period 3 is:

Low%	High%	Low$	High$	Total$
14%	86%	19.81	121.70	141.51

Figure 2.11: Payment Screen in Stage 2

Appendix 2.B Type classifications based on alternative criteria

Table 2.10 lists type-group classifications for eight criteria differing in test types (paired *t*-test or paired Wilcoxon signed-rank test), confidence levels (95% or 90%) and hypothesis symmetries (two tail or one tail). With only 40 observations, one may have concerns that the empirical distributions of the reports t-tests have yet to approach their asymptotic distributions. Rather than testing for normality of the distribution of paired differences, which would establish the validity of the *t*-tests, we report classifications based upon the nonparametric Wilcoxon signed-rank test. We find this generates qualitatively similar clusters. But we note that the Wilcoxon signed-rank tests offers lower power and results in slightly larger egoist clusters.

Criteria	*t*-test				*W*-test			
	$CL = 95\%$		$CL = 90\%$		$CL = 95\%$		$CL = 90\%$	
	2 Tail	1 Tail	2 Tail	1 Tail	2 Tail	1 Tail	2 Tail	1 Tail
Equity	24	28	28	31	26	31	31	33
Egoist	51	43	43	35	48	36	36	32
Efficiency	17	21	21	26	18	25	25	27

Note: This table provides the number of participants in each type based on eight classification criteria.

Table 2.10: Type classifications by alternative criteria

Notice that the results we reported in the Section 2.4 are robust to alternative type classification criteria. In unreported results, we ran the reduced form analyses for each of the three groups and for each of these type classification results from the Table 2.10. we find the estimated demand functions are similar to those reported in Table 2.2. Furthermore, we cannot reject the null hypothesis that the demand functions for Risk problems are the same across these three groups.

Appendix 2.C Additional materials for parametric analysis

Demand functions for CRRA utility function

With subjective expected utility specification and CRRA utility $u(x) = x^{1-\rho}/(1-\rho)$, each individual's preference is described by a curvature parameter ρ and a belief parameter α. The larger the ρ is, the more risk averse the individual will be. When ρ equals to 0, the individual will be risk neutral and when ρ equals to one, the power utility will degenerate to $\ln(x)$. The closer α is to one, the more objective the individual is perceiving the probability of the state.

The NLLS (and the general optimization) estimations are sensitive to the starting values of parameters. We calculate the starting values of α and ρ for each individual based on the interior optimal conditions of his constrained maximization problems. More specifically, we run the following OLS regression for a typical DM's (subjective) expected utility maximization.

$$\ln(x_i/y_i) = \beta_0^i + \beta_1^i \ln(q) + \varepsilon_i,$$

where ε_i is assumed to assumed to follow a normal distribution with zero mean and variance of σ_i^2. Based on the regressed parameters of $\hat{\beta}_0^i$ and $\hat{\beta}_1^i$, we can calculate the curvature parameter $\rho^i = 1/\hat{\beta}_1^i$ and the belief parameter $\alpha^i = \exp\left(-\hat{\beta}_0^i/\hat{\beta}_1^i\right)$. We also ran the OLS regression separately for $i \in \{\text{Risk, VoI}\}$.[35]

For this version of NLLS estimation with subjective expected utility model, the DM has the same belief parameter α for Risk and VoI problems. We use α_{Risk} as the starting values for the estimation algorithm. These estimation results are robust to alternative sets of starting values of α including α_{VoI} and various combinations of α_{Risk} and α_{VoI}.

Demand functions for uncensored CFGK and CKMS datasets

Notice the DM in CFGK and CKMS can choose allocations on the whole budget line, therefore, the optimization conditions are slightly different from those listed in subsection 6.2. With CRRA utility, we first similarly replace all boundary allocations of 0 with a small number such that the ratio between x and y is either ω or $1/\omega$. However, when there is no censoring constraint, the DM can choose boundary allocation $(x, y) = (0, z)$ or $(x, y) = (z/q, 0)$. For the first case, we replace 0 with ωz and for the second case,

[35] All parametric estimations at the individual level can be found in author's personal website.

we replace 0 with $\omega z/q$. Since ω is much smaller than q, this procedure will not affect the estimation. Solving the utility maximization problem yields the following optimization conditions:

$$\ln(x^*/y^*) = f\left[\ln(q), \omega; \alpha, \rho\right]$$

$$= \begin{cases} \ln(\omega) & \text{if} \quad \ln(\alpha) - \rho\ln(\omega) \leq \ln(q), \\ -\frac{1}{\rho}\left[\ln(q) - \ln(\alpha)\right] & \text{if} \quad \ln(\alpha) + \rho\ln(\omega) < \ln(q) < \ln(\alpha) - \rho\ln(\omega), \\ -\ln(\omega) & \text{if} \quad \ln(q) \leq \ln(\alpha) + \rho\ln(\omega), \end{cases}$$

where the first and third condition are derived from two corner solutions, one on the y-axis and the other on the x-axis; the second condition is similarly for the interior solution.

Demand functions for CARA utility

The utility specification for CARA is $u(x) = -\exp^{-Ax}$, where A is the absolute risk aversion. When $A = 0$, the individual will be risk neutral and the larger the A is, the more risk averse the individual will be.

Solving the utility maximization problem with CARA specification, we have following optimization conditions:

$$y^* - x^* = f\left[\ln(q), z; \alpha, A\right]$$

$$= \begin{cases} z & \text{if} \quad \ln(\alpha) + Az \leq \ln(q), \\ \frac{1}{A}\left[\ln(q) - \ln(\alpha)\right] & \text{if} \quad \ln(\alpha) < \ln(q) < \ln(\alpha) + Az, \\ 0 & \text{if} \quad \ln(q) \leq \ln(\alpha). \end{cases}$$

Similarly, three conditions correspond to a corner allocation, an interior optimal allocation and a censored allocation, respectively. We also estimate parameters A and α by minimizing the square sum of distances between true choices and theoretical demands for each subject:

$$\min_{(\alpha,A)} \sum_{j=1}^{n} \left[(y_j - x_j) - (y_j^* - x_j^*)\right]^2 = \sum_{j=1}^{n} \left[(y_j - x_j) - f(\ln(q_j), z; \alpha, A)\right]^2$$

The estimation procedures and the starting values are similarly calculated based on each individual's interior optimal conditions.

For the uncensored datasets in CFGK and CKMS, the optimal conditions for two expected utility models with CARA specification are:

$$y^* - x^* = f\left[\ln(q), z; \alpha, A\right]$$

$$= \begin{cases} z & \text{if} \quad \ln(\alpha) + Az \leq \ln(q), \\ \frac{1}{A}\left[\ln(q) - \ln(\alpha)\right] & \text{if} \quad \ln(\alpha) - Az/q < \ln(q) < \ln(\alpha) + Az, \\ -z/q & \text{if} \quad \ln(q) \leq \ln(\alpha) - Az/q. \end{cases}$$

Chapter 3

Revealed preferences over experts and quacks and failures of contingent reasoning

Abstract: In many economic scenarios, people face incomplete information about the payoff-relevant states of the world, and they may resort to different tests (e.g., analysts, medical diagnoses, or psychic octopuses) to obtain information to reduce their risk exposure. This chapter studies how people evaluate and choose tests. Are they able to avoid useless ones (quacks) and identify genuinely useful ones (experts)? Are they over-paying for quacks and under-paying for experts, and why? I develop a novel experiment wherein people face a rich and structured choice set of expert and quack tests and choose their favorite ones through a graphic coloring task. I find that people do fail to distinguish experts and quacks on a large scale, and they are over-paying for quacks but accurately paying for experts. These results are not driven by the standard explanations suggested in the literature, including belief updating bias, failure in best-responding, and intrinsic preference over certain information characteristics. Instead, I show that the main culprit is the failure of contingent reasoning in information processing. That is, people cannot correctly foresee how expert and quack tests influence their decision problems for all contingencies. The failure of contingent reasoning underlies many decision problems in behavioral economics and game theory and provide new implications for these fields.

3.1 Introduction

Imagine a decision-maker, DM hereafter, bets on the result of a race between two horses (L and R) to win a prize. The DM first checks the performances of two horses in the past 100 races, and she learns that horse L won 60 times, and R won 40 times. In addition, the DM can consult an analyst, get his recommendation of the bet, and then choose the bet. Suppose there is an analyst who correctly predicted 42 (70%) times among the 60 races in which horse L won and 18 (45%) times among the rest 40 races in which horse R won, should she solicit a recommendation from this analyst? When there are several such analysts, varying in prediction accuracies and competing with each other, to whom should the DM resort, and how much should she pay to get a prediction from the chosen analyst?

This question is common for many decision problems under incomplete information. For example, people choose among competing clinic doctors, diagnostic tests, financial advisors, and news sources to figure out the payoff relevant states of the world. I generically call them "*tests*".[1] They are not the decision-guiding signals per se but describe the process of how such signals are generated. This fundamental difference determines the DM evaluates tests before she actually acquires a piece of information — people have already chosen a horse analyst or a medical test when they receive race predictions or clinic diagnoses. If a test is ex-ante useless, I refer to it as a *quack*, and similarly, an *expert* test is ex-ante useful.

This paper studies how people choose and evaluate quack and expert tests. I first establish the theoretic framework for tests and the associated decision problems. The usefulness of a test, defined as the expected benefit for the decision problem from having or not having the test, is a joint output of decision problem-specific characteristics (e.g., prior information, test structure, and available actions), agent-specific characteristics (e.g., individual preferences and beliefs), and their interactions (e.g., reasoning process). Consider the particular analyst in the horse race example. If consulting him increases the probability to win the prize from the prior, he is an expert, and the probability increment measures his usefulness. Since the analyst's prediction is unknown at this point, a standard DM assesses the "expected" winning probability — a weighted average of the winning probability when conditioning on the analyst's prediction being l or r, respectively. Such a measure of test usefulness implicitly assumes that the DM anticipates all possible signals, correctly formulates posterior beliefs for each signal

[1] They are also called experiment (Blackwell (1951)), information structure (Green and Stokey (1978)), or information source in the literature.

and thus best-responds in bet choices, and more importantly, learns how does the test structurally interact with beliefs and optimal bet choices.

The above evaluation process implies four major channels leading to the failure in distinguishing quack and expert tests. They include (1) the DM is unable in *updating beliefs* as a Bayesian; (2) she makes *sub-optimal choices* given her beliefs; (3) she has an *intrinsic preference* over certain types of tests; and (4) she lacks *contingent reasoning* in how the value of a test changes with its influence on beliefs and optimal actions.

Using a novel graphic experiment, I elicit people's preferences over tests and examine how they are explained by the aforementioned channels. Individuals face fourteen decision problems, each analogous to the horse race example, winning a prize through a bet on state L or R. Before the bet decision, they see a coloring representation of a set of tests, and they move sliders to control color compositions and select their favorite tests. The choice set in each problem consists of both expert and quack tests. They are on a common linear budget with two accuracies as two commodities, reflecting trade-offs between receiving news confirming one state versus the other. I construct fourteen budgets with multiple prices and expenditure levels. They generate rich variations both within and across the choice set of tests, and thus I can identify different decision rules. Given the chosen test, subjects also estimate the likelihood of two states and make bet choices. These auxiliary tasks are part of the incentive schemes of choosing an instrumentally valuable test, and on the other hand they elicit people's posterior beliefs and strategies for the evaluation of tests. Altogether, the experiment is user-friendly and provide diagnoses of different mechanisms.

I find people fail to distinguish experts and quacks on a large scale. Subjects frequently select useless tests at the aggregate, decision problem-wise, and individual level. The failure is *not* driven by belief updating bias, best-responding bias, or intrinsic preferences over certain types of tests. Subjects' reported posteriors are very close to Bayesian posteriors, and they rarely choose sub-optimal bets. Moreover, none of the variables measuring belief bias or action sub-optimality explains quack versus expert choices. I also construct test-specific and posterior-specific measures to capture subjects' preferences over several test characteristics. I find they have similar distributions between the group of experts and quacks, excluding intrinsic preferences as a channel for quack choices.

I also find that people are over-paying for quacks but accurately paying for experts. In particular, given the chosen test is an expert, subjects correctly identify the most useful expert; while given a quack test, they choose the most distant quack (the quack test with one accuracy equals to either zero or one). Based on subjects' descriptions of their decision processes, I identify three simple rules employed to evaluate tests and further establish their roles in causing such a pattern. These rules reflect subjects' reasoning in resolving overall uncertainties (*entropy-reducing rule*), discriminating processes of evidence generation (*evidence-separating rule*), and polarizing the chances of two signals (*signal-separating rule*). All of them justify "extreme" tests on a budget, leading to the pattern of choosing either the optimal expert or the distant quack. However, they cannot fully rationalize the choice of quacks versus experts.

The above two findings confirm a universal bias in reasoning when a test is useful. I refer to the bias as *the failure of contingent reasoning in information processing*. It occurs when subjects fail to recognize that the value of a test depends on its influence on the associated decision problem. This bias hinges on a simple intuition: when a test induces posteriors support the same action for different signals, the test is a quack; otherwise, it is an expert.[2] In the horse race example, the DM bets horse L under the prior belief that horse L has a 60% chance to win the prize. When faced with the analyst with the accuracy pair (70%, 45%), a Bayesian DM's posterior belief is 66% over state L after a signal l and 50% after r. Both posteriors support the same bet on horse L, indicating this particular analyst is a quack. From the ex-ante perspective, the DM's chance to win the prize remains the same as the prior due to the law of iterated expectations.

This paper contributes to a growing literature on preferences over information structures. A bulk of work consider the case when information is non-instrumental. Falk and Zimmermann (2016), Ganguly and Tasoff (2017), and Nielsen (2018) study the intrinsic preference over test informativeness through preferences over the timing of information and the resulting resolution procedures of uncertainty. They find people typically prefer to receive the decision-irrelevant information sooner.[3] Masatlioglu, Orhun and Raymond (2017) provide evidence for an intrinsic preference over positively-skewed information structures. In the setting when information is instrumentally valuable, Ambuehl and Li (2018) find that people undervalue expert tests due to belief updating biases. Charness, Oprea and Yuksel (2018) and Montanari and Nunnari (2019) study how people choose between prior-confirming and contradicting information structures, and they find people prefer confirming ones at the cost of informativeness. I consider a framework for both instrumentally useless (quacks) and useful (experts) information structures and study whether people can differentiate them. Moreover,

[2] Early work such as Hirshleifer (1971) has already noticed the insight that information can be useless if it does not impact choices. Recently, Lara and Gossner (2020) exploit the bilinear duality structure between payoffs and beliefs to study the value of information in the same spirit.

[3] Theoretical work on the preferences over early versus late resolution and gradual versus one-shot information include Kreps and Porteus (1978), Loewenstein (1987), Epstein and Zin (1989), Grant, Kajii and Polak (1998), Dillenberger (2010), and Ely, Frankel and Kamenica (2015).

I focus on people's reasoning biases that lead to failures in identifying experts and quacks.

This paper also contributes to the literature on the failure of contingent reasoning. It often describes people's inability in recognizing the optimal actions for all possible contingent states or their opponents' actions.[4] An agent without contingent reasoning ability cannot identify dominant strategies. Therefore, many studies on contingent reasoning are the same as on the empirical validity of the dominant strategy in different games and mechanisms. Tversky and Shafir (1992) make subjects play one-shot prisoner's dilemma games and find people have trouble realizing that their dominant strategy does not rely on opponents' choices. Cason and Plott (2014) find people do not report truthfully (even though it is a dominate strategy) under the Becker, DeGroot and Marschak (1964) (BDM) elicitation mechanism. Harstad (2000) finds people fail to identify the dominant strategy in the sealed-bid second-price auction even after sufficient learning opportunities. Esponda and Vespa (2014) show similar mistakes in a common-value voting experiment. Chen (2008) reviews such evidence for the context of public good provision.

Esponda and Vespa (2019) generalize the state space for strategic games and formally define the failure of contingent thinking as violating Savage (1972)'s sure-thing principle. They also experimentally show such failures are robust across several decision problems and games. In this paper, I identify a different form of contingent reasoning. The contingencies are on information structures, not on states. More precisely, the state-contingent failures occur since the DM is not partitioning the states between those where her choice does matter and those where it does not. While the test-contingent failures occur since the DM is not partitioning the information structures between those with which her optimal strategies are pooling across signals and those with which are separating. Such failures have a significant implication for how people seek out information sources, how data sellers provide information products, and the general information design problems.

The rest of the paper is structured as follows. Section 3.2 introduces the theoretical framework for expert and quack tests and their values. Section 3.3 describes the experiment to elicit preferences over tests. In Section 3.4, I present the experimental results on failures in distinguishing and evaluating experts and quacks. Section 3.5 and Section 3.6 examine the mechanisms behind these failures, and Section 3.7 discusses extensions and implications of my findings. Section 3.8 concludes.

[4] Martínez-Marquina, Niederle and Vespa (2019) consider the contingent reasoning implied in the strategic requirement of subgame perfect Nash equilibrium. That is, people cannot correctly foresee how others behave for all contingencies, no matter they will arise or not.

3.2 Theoretic framework for expert and quack tests

3.2.1 Setup: states, signals, and tests

The DM faces a decision problem with two states of the world $\omega \in \{L, R\}$ and two actions $a \in \{L, R\}$. Each action corresponds to a bet on the state, yielding a payoff of π if it matches the true state and zero otherwise. I denote the objective prior belief of state L by a scalar μ and assume $\mu \in [1/2, 1)$. The DM may also resort to a test for a signal $s \in \{l, r\}$. The signal is potentially informative for the inference of the true state. Each test governs a generation process of signals, and the DM chooses a test before knowing the signal to be generated. By analogy to the horse race example, I define a test as a pair of two state-contingent probabilities (accuracies) and denote it by (p, q), in which $p \equiv \mathbb{P}(s = l \mid \omega = L)$ and $q \equiv \mathbb{P}(s = r \mid \omega = R)$. I consider a collection of tests satisfying $p \geq 1 - q$. That is, the probability of signal l under state L is at least as large as that under state R. I call such tests *admissible*, and they are in the set denoted by $\mathcal{T} \equiv \{(p, q) \mid p + q \geq 1,\ 0 \leq p, q \leq 1\}$.[5]

How to choose the optimal bet for the decision problem? I assume the DM is probabilistically sophisticated as in Machina and Schmeidler (1992). In my setup, it simply says that the DM's preference is consistent with her beliefs. Therefore, her optimal action is to bet the state she thinks has at least a fifty percent chance to occur. This assumption admits the DM being a non-expected utility maximizer. For example, the DM may transform her belief probabilities with a strictly increasing function as in Tversky and Kahneman (1979) and Tversky and Kahneman (1992) (also see Wakker (2010)). The threshold for the belief is fixed at one half, under which the DM is indifferent between two bets. This threshold is invariant of the information environment characterized by the prior, the signal, and the test and is exclusively determined by the symmetric payoff structure of two bets. Doing so removes the influence of the preference over outcomes and leaves all the analyses within the probability space. The DM's utility over outcomes may be linear, concave, or convex, but she always chooses the optimal bet by comparing her belief of each state with the threshold. In section 3.7, I discuss how my setup can be easily extended to decision problems with alternative payoff structures.

A test affects the optimal bet by inducing a distribution of posterior beliefs over the state. Given a prior μ, I denote the posterior induced by a particular test (p, q) by a scalar $\mu_s(p, q; \mu)$, which is the DM's belief of state L after signal s. Before knowing

[5] This set partitions the test space $[0, 1] \times [0, 1]$ into halves and is rich in capturing all possible posterior distributions. It is symmetric to the tests in the other half of the partition, and all theoretic analyses remain the same by swapping two signals.

the signal, the DM treats the posterior as a random variable with two realizations $\{\mu_l, \mu_r\}$. The probability mass of each realization is the unconditional probability of the corresponding signal $\mathbb{P}(s)$. Provided the signal s, the DM's optimal action is to bet state L whenever the posterior belief after the signal is larger than one half and to bet state R otherwise, and thus she believes that her chance to win the prize is $\max\{\mu_s, 1-\mu_s\}$. Since the DM does not know whether the signal is l or r yet, she takes expectation of these two signal-contingent winning chances. Let $v(p,q;\mu)$ denotes the ex-ante expectation of the probability to win the prize, we have

$$v(p,q;\mu) = \max\{\mu_l, 1-\mu_l\}\,\mathbb{P}(s=l) + \max\{\mu_r, 1-\mu_r\}\,\mathbb{P}(s=r).^{6}$$

The defined winning chance provides a criterion to evaluate and compare different tests. When there is no test, the DM bets L and expects to guess the state correctly with a chance equals to the prior. The difference between $v(p,q;\mu)$ and prior μ measures how useful the test is to the decision problem from having or not having the test. A quack test leads to a zero increment in the chance to win the prize, and an expert test leads to a strictly positive one. The increment summaries how much extra "certainty", compared to that under prior knowledge, the DM gains from the test. For a given prior, I simply define the *value of a test* as the expected winning probability $v(p,q;\mu)$. Notice that the DM's interim decision-making elements, precisely, her belief updating process and action best-responding, root in this measure. When there are several tests available, the DM should choose a test she thinks has the largest increment in the expected winning probability. Equivalently, she is selecting a distribution over posterior beliefs that guides her bet choices the most.

3.2.2 Distinguishing experts and quacks in a rational benchmark

Consider the benchmark of a *rational* DM who updates as a Bayesian and best responds to her Bayesian posteriors. Due to the martingale property of Bayesian updating, any arbitrary test will induce a mean-preserving posterior spread of the prior. Moreover, if a test is admissible, its Bayesian posterior of state L after signal l is no smaller than that after signal r. The vice versa also holds. Proposition 3.1 formalizes the idea.

[6] Notice that posterior μ_l and μ_r are mappings from the admissible set $\mathcal{T}$ to $[0,1]$. The unconditional probabilities of signal l and r also depend on the test (p,q) and the prior μ.

Proposition 3.1. *An admissible test induces a Bayesian posterior spread such that observing signal l (or r) increases the posterior of the state L (or R) relative to the prior, and vice versa. Mathematically, for any prior* μ*, a test* $(p,q) \in \mathcal{T} \iff \mu_r^{Bayes}(p,q;\mu) \leq \mu \leq \mu_l^{Bayes}(p,q;\mu)$.

A rational DM will bet state L if the signal realization is l. Given the prior is in favor of state L ($\mu \geq 1/2$), this strategy is consistent with her posterior belief over state L being larger than one half.[7] When the signal is r, she compares the corresponding Bayesian posterior with the threshold of one half. If it is larger than one half, she bets state L and wins the prize with a chance of μ_r^{Bayes}; otherwise, she bets R and has a chance of $1 - \mu_r^{Bayes}$. Taking expectations of these two contingencies, a rational DM's ex-ante winning probability of the prize is

$$v^{Bayes}(p,q;\mu) = \mu_l^{Bayes}\mathbb{P}(s=l) + \max\left\{\mu_r^{Bayes}, 1-\mu_r^{Bayes}\right\}\mathbb{P}(s=r).$$

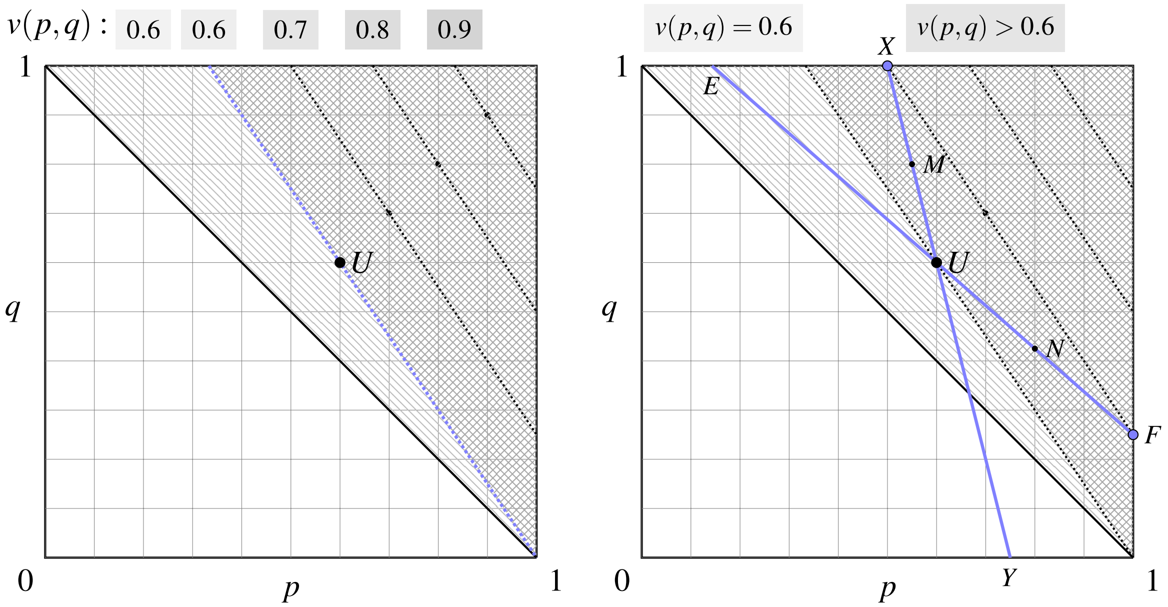
Whether a test is an expert or a quack rests on μ_r^{Bayes}, the induced posterior belief after signal r. If it is smaller than one half, the test is an expert; otherwise, the test is a quack. Intuitively, a test is useful only if it induces posteriors that support alternative optimal actions for the decision problem, instead of the same one indicated by the prior. Proposition 3.2 states the condition that distinguishes expert and quack tests. Moreover, it shows the value of each expert test is a linear combination of two state-specific accuracies, with prior beliefs as their weights. This assessment is empirically convincing. When an expert is not able to predict the true state for sure, people evaluate his expertise based on a weighted average of his prediction accuracy for each state.

Proposition 3.2. *Given a rational DM and a prior* μ*, an admissible test* (p,q) *is an expert (a quack) if and only if the condition* $(1-p)\mu - q(1-\mu) < 0$ *is (not) satisfied. The value of an expert test is* $v^{Bayes}(p,q;\mu) = p\mu + q(1-\mu)$*, and the value of a quack test is* μ.

Figure 3.1a plots the partition of experts and quacks. The prior is three fifths. Each point lying above the diagonal line is an admissible test, with the x-axis coordinate as the accuracy for state L and that on y-axis for state R. The parallel dotted lines are a rational DM's indifference curves over experts' accuracies p and q. They share the same slope $\mu/(1-\mu)$, which is the prior odds of state L versus R. One indifference curve (blue) partitions the admissible space into two sets, one corresponds to quack tests

[7] I assume the DM will choose action L when her posterior beliefs over two states are the same. Under my construction of priors and admissible tests, the indifference scenarios with a Bayesian posterior of $1/2$ after signal l only occurs when the prior is $1/2$, and the test satisfies $p+q=1$. Notice the Bayesian posterior after signal r also equals to $1/2$ for these scenarios. Therefore, all analyses will not be affected by this indifference-resolving procedure.

(stripped area) and the other one to experts (crosshatched area). For an arbitrary prior μ, the partitioning indifference curve is characterized by $q = \frac{\mu}{1-\mu}(1-p)$ and has two fixed points (μ, μ) and $(1, 0)$. The partition line under a larger prior will pivot around point $(1, 0)$ towards northeast, generating a larger set for quack tests. It implies that when the prior is already very informative for the state, more tests are quacks as it is more difficult to improve the chance to win the prize.



(a) Partition of expert and quack tests

(b) Test choices on linear budgets

Figure 3.1: (a) A rational DM's indifference curves and partitions of experts and quacks; (b) an example of two linear budgets used to elicit preferences over tests. The prior is fixed at three fifths. The x-axis is the conditional probability of signal l under state L, and the y-axis is the conditional probability of signal r under state R. The numbers on top are the values of tests on depicted indifference curves.

3.2.3 Eliciting preference over tests via linear budget sets

In practice, people's indifference curves over tests may be non-linear. For instance, if a DM has a preference for tests with a large difference between two accuracies, her indifference curves will be convex. However, all indifference curves are downward sloping regardless of their shapes.[8] They reflect how people make trade-offs between two state-contingent accuracies. To compare two tests, one with a higher p and the other with a higher q, the DM is trading off the probabilities to receive good news from

[8] The downward sloping indifference curves for experts implicitly assume the preference over test accuracies are monotonic and non-satiated.i

state L versus R.[9] In other words, people's preferences over tests are revealed through their choices of the accuracy bundle (p,q) when faced with trade-offs between p and q. I follow this idea and elicit people's preferences over tests via linear budgets $p+mq=z$, wherein m is the relative price, and z is the accuracy expenditure. Each linear budget is a choice set of tests, and the DM selects her most preferred alternative, knowing that each one percent increase in accuracy q implies a decrease of m percent in accuracy p.[10]

Figure 3.1b illustrates two linear budgets, XY and EF. The former budget is steeper than the indifference curves of a benchmark DM, and the latter one is flatter. They interact at the symmetric quack test U. Such construction ensures that U divides both budgets into two segments, one for experts and one for quacks. Meanwhile, experts are lying on the upper segment for budget XY and on the bottom segment for budget EF. When the slope of the budget is larger than that of indifference curves, like in the case of XY, the commodity q is relatively cheaper. Therefore, it is optimal to choose the accuracy bundle X with $q=1$. Similarly, the commodity p is relatively cheaper on EF, making F the most useful expert. I examine whether people can distinguish experts and quacks by the location of the chosen test being on the expert segment or not. The deviation of the chosen test from the optimal one, on the other hand, investigates whether people are over-paying for quacks and under-paying for experts.

I choose particular pairs of budgets such that for each expert on one budget, there is a unique test on the paired budget having the same value. For instance, consider test M and N. They are equally valuable but skew to different state-specific accuracies. They also induce posteriors differing in spreads, skewness, or entropy. With such budget pairs, I can identify people's intrinsic preferences over certain test characteristics and examine the associated decision rules. I can also study the relationships between people's intrinsic preference and preference over tests for their instrumental value.[11]

The linear budget approach is widely used in experimental literature to measure people's preferences in various settings, for instance, their altruism preferences under

[9] When there are no trade-offs between two tests, one must be at least as accurate as the other one for both states. The value for the former one is no smaller than the latter. In particular, the former one is Blackwell more informative than the latter when they are experts.

[10] Equivalent formulations include the trade-offs between a type I and type II statistical error or between the false positive rate and true positive rate. The later one is also the Receiver Operating Characteristic (ROC) curve(Hanley and McNeil (1982)), which is extensively used for the validity of a diagnostic test in clinical decisions. Chan, Gentzkow and Yu (2019) empirically estimates the radiologists' skills based on their diagnosis accuracies under the ROC framework. My study provides experimental evidence on the determinants of ROCs through a linear approximation.

[11] Subjects in my experiment can choose on a full budget. Since steep budgets intersect the diagonal line, the tests below the diagonal line are not admissible. However, all theoretical analyses are still valid. For such choices, receiving signal l (r) will increase the posterior of state R (L) from the prior. Since the state of L is advantageous in prior, all of these non-admissible tests are quack tests

the dictator game (Andreoni and Miller (2002)), social preferences over efficiency and equity behind the veil of ignorance (Heufer, Shachat and Xu (2019)), risk preferences (Choi et al. (2007*a*), Choi et al. (2014)) over two state-contingent assets, and time preferences (Andreoni and Sprenger (2012)) over current and future payments. With a rich choice set and an intuitive representation of trade-off problems, this approach allows for robust inferences from the elicited preferences. In addition, it provides a framework to apply the classic nonparametric revealed preference techniques (see Afriat (1967) and Varian (1982, 1983)) to examine whether preferences are standard.

3.3 Experimental design

3.3.1 Tasks

The experiment consists of fourteen decision problems. In each problem, subjects choose their most preferred test from a budget set in order to bet the state correctly. Table 3.1 lists the priors and budget specifications for all problems in the experiment. They show up in a random order. Each row gives a budget pair consisting of a steep and a flat budget, and they interact at a symmetric test, just like budget XY and EF show in figure 3.1b. I refer to the interaction point as a “Pivot” point. For each pair in the top five rows, pivot points are quack tests that coincide with priors. For budget pairs on the bottom two rows, the pivot is a fixed expert test $(3/4, 3/4)$. Figure 3.8 in Appendix 3.B depicts all budgets in the accuracy space.

No.	Prior	Steep budget	Flat budget	Pivot
P1	$1/2$	$2p+q=3/2$	$p/2+q=3/4$	$1/2$
P2	$3/5$	$2p+q=9/5$	$5p/4+q=27/20$	$3/5$
P3	$3/5$	$4p+q=3$	$7p/8+q=9/8$	$3/5$
P4	$2/3$	$4p+q=10/3$	$3p/2+q=5/3$	$2/3$
P5	$3/4$	$6p+q=21/4$	$5p/2+q=21/8$	$3/4$
P6	$3/5$	$6p+q=21/4$	$5p/2+q=21/8$	$3/4$
P7	$2/3$	$6p+q=21/4$	$5p/2+q=21/8$	$3/4$

Table 3.1: Fourteen budgets used in the experiment

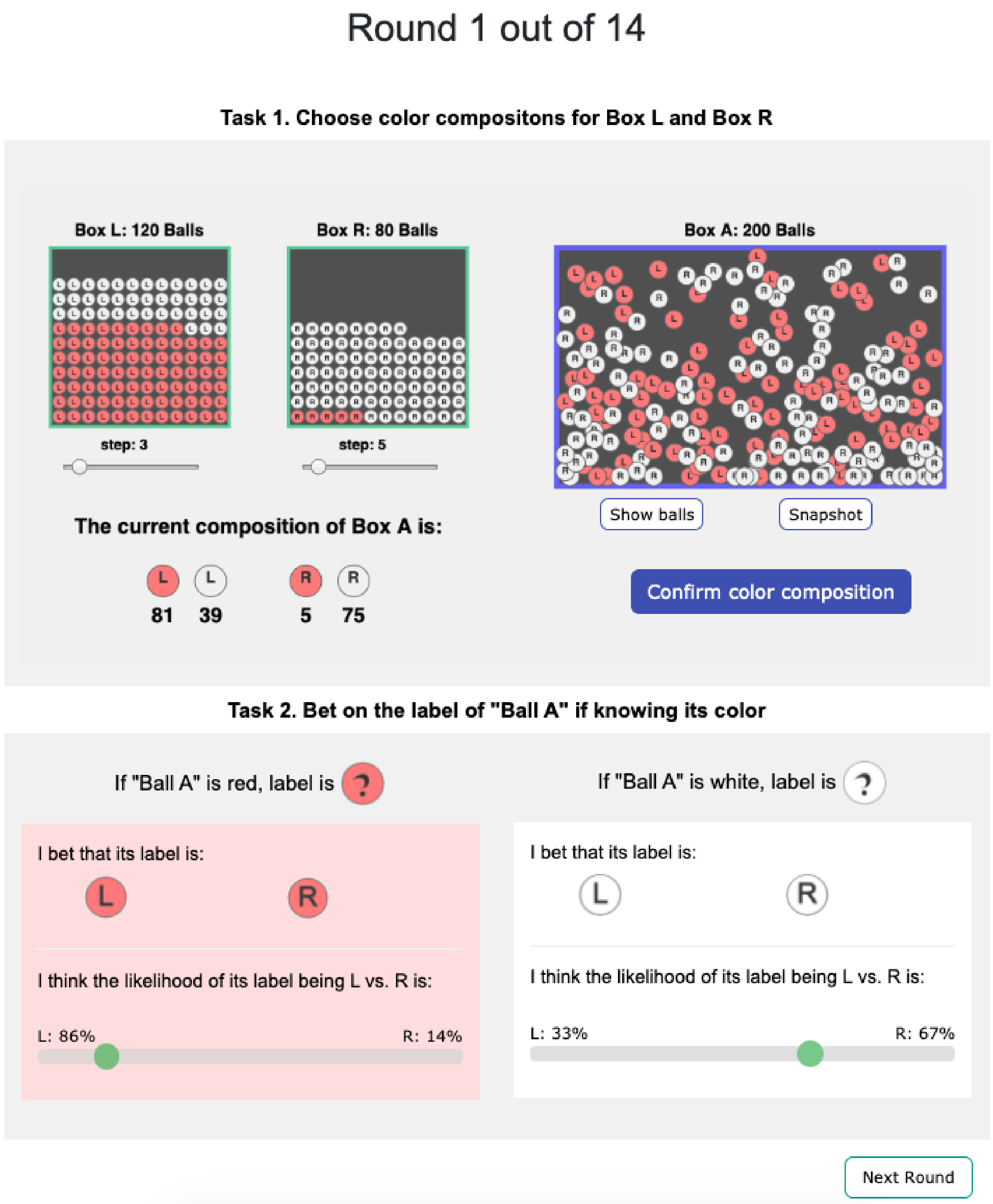
I develop a new interface to implement the choice of tests. Instead of making subjects choose on an abstract budget as in the literature, I present the decision problem as a graphic coloring task. Figure 3.2 shows the screenshot for a typical choice round. Box L and R at the top-left corner contain some labeled balls. Later all balls will be filled into

Box A, and one ball (called Ball A) will be randomly drawn from it. A participant wins ten pounds if correctly betting the label of Ball A. The setting resembles the horse race example. The proportion of L balls in Box A indicates the prior probability of state L. Before putting all balls into Box A, the participant can color some balls in Box L and R by two sliders below the boxes. Moving a slider to the right (left) will increase the number of red (white) balls in the corresponding box. The step for the increments for each box is indicated above the slider. The initial locations of two sliders are determined by the pivot point, reflecting a default option of a symmetric test.

Choosing a color composition for two boxes is equivalent to choosing a test on a linear budget. First notice the number of balls in Box L and R imply the prior belief of each state. The red ball serves a signal realization of l, and the white ball is r. The proportion of red balls in Box L is accuracy p, and the proportion of white balls in Box R is q. The trade-offs between two accuracies are achieved by linking two sliders such that the colored balls in box L and R increase (decrease) with each other. For instance, in this particular round, every time the participant adds three red balls for Box L (p increases), five red balls will be automatically added to Box R (q decreases). More details about the construction of two boxes and their steps for each budget can be found in Table 3.11 in Appendix 3.B.

The top-right corner of the interface shows what Box A looks like for the chosen color composition. When subjects move sliders, Box A's coloring changes accordingly. With such interaction, it is clear to subjects that the color choices are instrumentally valuable. The state is not realized yet and their color choices matter for the correct guess of the state. I also provide an animation with the colored and labeled balls bumping with each other. It visualizes the random process of signals and states, and thus will remove subjects' suspicions about how uncertainties are resolved and how their payoffs are determined in the experiment.

After choosing the color composition, participants provide likelihood estimates for the true state and their bet choices conditional on each one of the signal realizations. They are shown as Task 2 at the bottom of the interface. These auxiliary answers provide direct evidence on the mechanisms underlying people's preferences over tests. In particular, the likelihood estimates measure people's posterior beliefs, and the bet choices indicate whether people are best-responding given their beliefs. Notice that they are not separated from task of choosing tests. They determine the process to resolve the signal and the state and play an important role in the incentives of the test choice. If Ball A turns out to be red, the participant will be rewarded for her estimates and bet choices conditioning on the signal of red.



Round 1 out of 14

Task 1. Choose color compositons for Box L and Box R

Box L: 120 Balls

step: 3

Box R: 80 Balls

step: 5

Box A: 200 Balls

The current composition of Box A is:

L	L	R	R
81	39	5	75

Task 2. Bet on the label of "Ball A" if knowing its color

If "Ball A" is red, label is ?

I bet that its label is: L R

I think the likelihood of its label being L vs. R is:

L: 86% R: 14%

If "Ball A" is white, label is ?

I bet that its label is: L R

I think the likelihood of its label being L vs. R is:

L: 33% R: 67%

Figure 3.2: Experimental interface for a typical decision round

I present the belief updating task with natural frequencies and provide additional incentives for Bayesian posteriors. Given a color composition, the task of providing likelihood estimations is very similar to Grether (1980)'s "urn-and-ball" experiment. However, I avoid all descriptions of probabilities and present states and signals in frequency formats. For example, subjects can check the numeric composition of each type of balls in Box *A* when formulating posteriors. The frequency format describes a direct and convincing process about how the signal is naturally sampled and how the random state is resolved. Gigerenzer and Hoffrage (1995) show that the natural frequency formats substantially improve (up to 50%) Bayesian inferences than probability formats in several thousand Bayesian problems. Also, subjects are rewarded for being close to a Bayesian. The participant is told that a mathematician is facing the same color composition and also providing an estimate for the state. If the absolute difference between her estimate and the mathematician's is smaller than five percent, she will earn a bonus of one pound and a half; if otherwise, the absolute difference is lower than fifteen percent, the bonus is fifty cents. This incentive scheme is straightforward to subjects.

3.3.2 Procedures

I ran the experiment in March 2020. After running the pilot session in the lab, the campus is closed. I recruited 64 subjects on Prolific platform and ran the experiment online.[12] The average payment was £7.25, not including a show-up fee of £4. Average duration of the experiment was 45 minutes.

Before the main task, subjects need to read the instructions (see Appendix 3.D) carefully and complete a quiz. The quiz questions are not trivial. On average, subjects spent 18 minutes on instructions and quiz questions, and it is similar to that of the pilot lab session.

After the experiment, one problem is randomly selected for payment. Subjects see a summary screen (see Appendix 3.D) showing all information relevant for their payoffs: the random round selected for payment, the color and the label of ball *A*, their coloring choices, likelihood estimations and bets in that round, and the estimations from a Bayesian mathematician. I ask them to share thoughts about how they chose the color composition and the bets.

The final part of the experiment is a questionnaire that includes questions regarding demographic variables, self-evaluated psychological attitudes, and measures of cognitive reasoning. I consider three commonly used psychometric tests: Frederick (2005)'s cognitive reflection test (CRT) for the tendency to use heuristics, Wason (1968)'s

[12] To mimic the lab session, I imposed the pre-screening conditions including the age, undergraduate or master student, English proficiency, ect. In general, their choices are comparable with the those from the pilot session in the lab.

selection task for the measurement of deductive reasoning, and logic-based Syllogism questions.

3.4 Experimental results

This section reports experimental results on subjects' choices of tests. In particular, I examine whether they can distinguish expert and quack tests on linear budgets and whether they are choosing the most useful experts. Figure 3.3 provides an array of scatter plots for four subjects' choices and budget sets. Subject 15 and 50's choices, in the top row of the array, exhibit a taste for tests with intermediate accuracies. However, most of Subject 15's choices are on the expert segments, while Subject 50 chooses tests on the quack segments for over half of the budgets. Subject 22 and 41, whose choices are in the bottom row, prefer tests on the border. Subject 22 is able to recognize the useful one between two border tests, and Subject 41 is not. Overall, subjects are heterogeneous in their preferences over tests, but they frequently choose quacks.

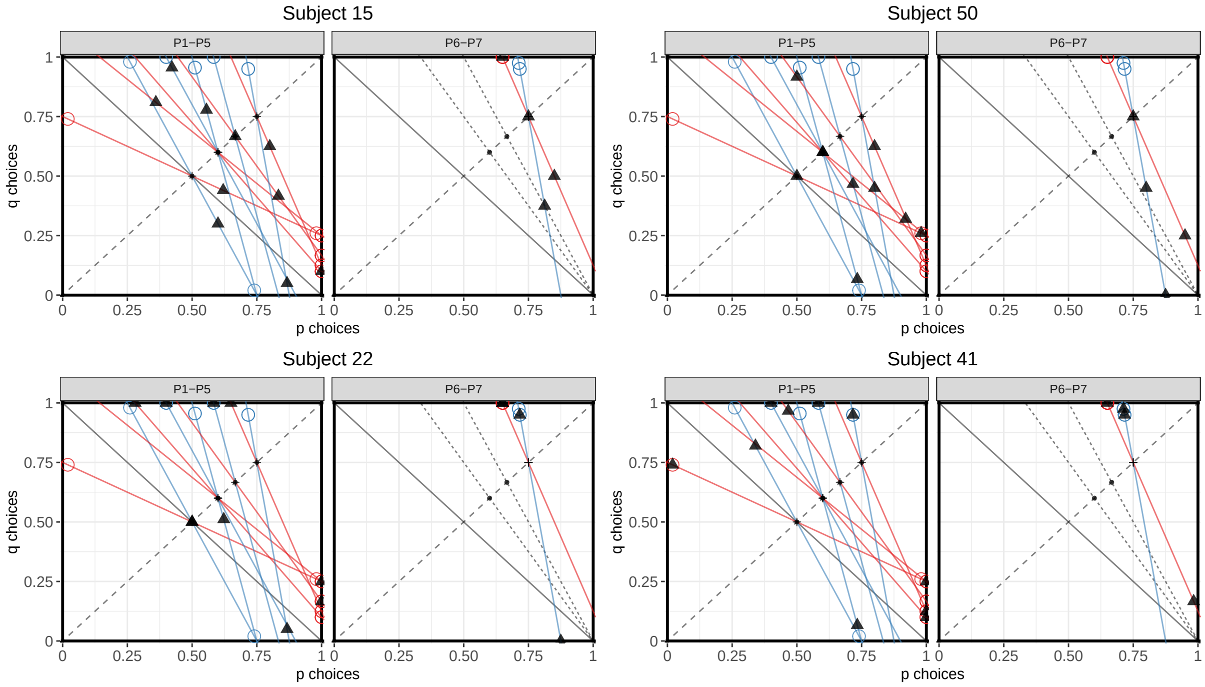


Figure 3.3: Illustrative examples of actual test choices and optimal ones. The budgets for P1-P5 and P6-P7 are shown in separate panels. The steep budgets are in red and the flat ones are in blue. Solid triangles show the subject's actual choices of (p, q), and hollow circles show the optimal ones.

I proceed by providing statistical examinations of subjects' choices at the aggregate, decision problem-wise, and individual level. These results demonstrate that subjects fail to distinguish expert and quack tests on a large scale.

3.4.1 The failure in distinguishing experts and quacks

At the aggregate level, subjects frequently choose quack tests. Table 3.2 reports the frequencies of quacks and experts across different budget pairs. The frequency of quack choices is 35% for all decisions, and it is significantly different from the average chance if randomly selecting tests on the same set of budgets (41%). Around 20% of subjects' choices are on the quack segments of budgets in P1, P6, and P7. These budget pairs on average admit fewer quack tests.[13] If we exclude them, subjects choose quack tests 47% of the time, and it is again significantly different from the random chance (62%) for budgets in P2-P5.

	Quack	Expert	Total
P1-P7	286 (35%)	526 (65%)	812
P1	24 (21%)	92 (79%)	116
P2-P5	219 (47%)	245 (53%)	464
P6-P7	43 (19%)	189 (81%)	232

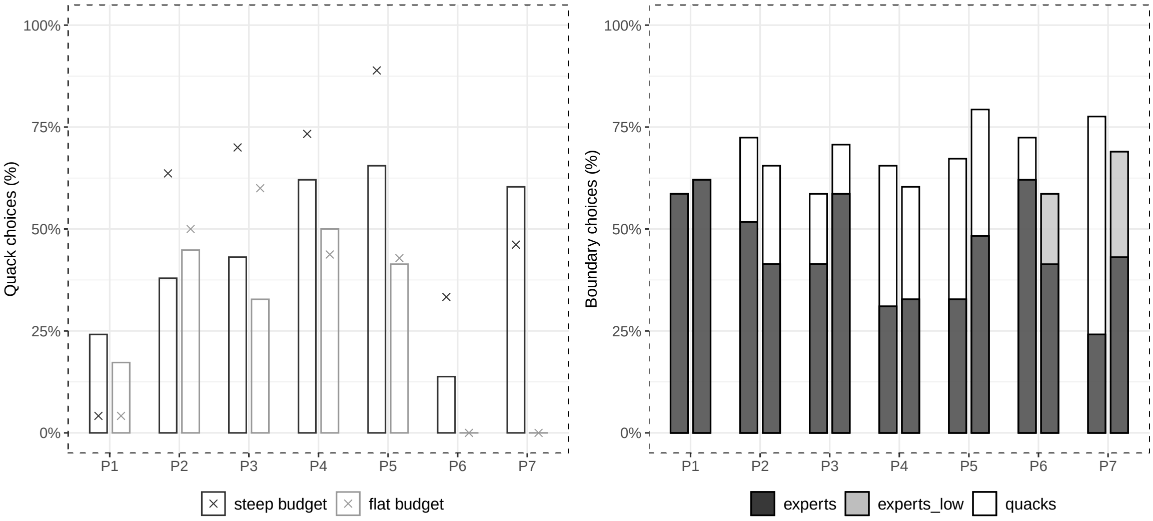
Table 3.2: Frequencies of quack and expert choices

Figure 3.4a plots the average rate at which subjects choose quack tests for each budget. Though the rate varies, subjects frequently choose quacks for almost all budgets. Nonetheless, they are not choosing randomly. They still identify experts more often than a random DM for most of the budgets.[14] In general, steep budgets induce more quack choices than flat ones.[15] However, a steep budget has a larger segment of quack tests than a flat budget, and therefore is easier for subjects to make mistakes in choosing quacks. After controlling the random chance of each budget, two types of budgets are not significantly different in predicting quack choices.

[13] For instance, almost all tests (except the symmetric one) on budgets in P1 are experts. There is no quack test for the flat budgets in P6 and P7.

[14] For ten out of fourteen budgets, the proportion of quack choices is significantly different from the corresponding random chance. The four budgets with insignificant results are the flat budget in P2 and P5 and two budgets in P4.

[15] The test is based on a probit regression with the binary quack choices as dependent variables and the binary budget type as independent variables. The estimate shows that steep budgets have 25% more chances in predicting quack choices than flat budgets. The difference is significantly different from zero, with a p-value of 0.0325.



(a) Frequency of quack choices

(b) Frequency of border choices

Figure 3.4: The histogram of quack choices and border choices for fourteen budgets. Points with cross markers in the left panel are a random DM's frequencies of selecting quack tests. The category of "expert-low" in the right panel indicates the non-optimal border choices for the flat budgets in P6 and P7. They are the least useful expert tests for these budgets.

The quack choices are also pervasive among all subjects in the experiment. They select quack tests for around five out of fourteen decision problems. Each participant chooses quack tests at least twice, and over half of them choose four times. Subjects on average still beat a random DM in recognizing expert tests. Individual frequencies of quack tests are significantly different from the random chance of selecting a quack test.[16]

3.4.2 The failure in evaluating experts and quacks

I continue to examine what kind of quack and expert tests subjects choose. Are the quack choices consistent with specific decision patterns or simply driven by trembling hands? Are the chosen expert tests the most useful ones? I define a participant's evaluation of a test as its distance deviation from the optimal test on the same budget. This deviation measures the opportunity cost of choosing the test by giving up the optimal one. A substantial deviation means that the subject is over-paying for the test.

[16] The p-value is 0.0002 for the Wilcoxon rank-sum test and 0.0008 for the t-test. If not specified, all tests reported in this section are based on t-test, Wilcoxon rank-sum test for unpaired variables, Wilcoxon signed-rank test for paired variables, or proportion test. I will not report p-value if it is smaller than 0.0001.

Since each budget differs in slopes and lengths, I normalize the deviation as a fraction of budget length. A deviation of zero means the choice itself is the optimal test, and a value of one implies the test is on the non-optimal endpoint of the budget.

The average deviation across all subjects and all budgets is 0.41, suggesting that subjects' choices are far from the most useful expert tests. However, its distribution uncovers a significant polarization in test choices. 41% of choices have zero deviations, and 26% have deviations of one. This result means that subjects overwhelmingly prefer tests that are endpoints of a budget. I call them border tests. When such test choices are on the optimal border, subjects are choosing the most useful experts; otherwise, they are choosing the most distant quacks.

Table 3.3 further summarizes three categories of subjects' choices: border, center, and interior test. Aside from border choices, subjects choose center tests which have symmetric accuracies 14% of the time. The rest 19% of choices are categorized as interior tests. The bottom two rows of the table show the test categories for expert and quack choices. Over half of the subjects' choices are on the border, regardless of their usefulness. Furthermore, 72% of border tests are the most useful expert tests, and this proportion is significantly different from the chance of one half.

	Border	Center	Interior	Total
Pool	544 (67%)	113 (14%)	155 (19%)	812
Quack	154 (54%)	78 (27%)	54 (19%)	286
Expert	390 (74%)	35 (7%)	101 (19%)	526

Table 3.3: Frequencies of three categories of test choices

The category results at both the decision problem and individual level further confirm the findings. Figure 3.4b plots the proportion of border choices and its compositions of experts and quacks for each budget. Subjects uniformly prefer border tests, and within the category, their choices are more likely to be expert tests for most of the budgets.[17] At the individual level, subjects choose border tests nine times on average, and over one third of subjects choose border tests at least twelve times.

What are the consequences of choosing quacks or non-optimal expert tests? I compare the expected Bayesian winning probabilities of subjects' actual choices and those of the optimal tests. Subjects' test choices entail a chance of 0.684 to win the prize on average, which is significantly smaller than the one if they choose optimally (0.720).

[17] For nine out of fourteen budgets, the proportion of expert choices within the border category is significantly different from the chance of one half. The five budgets with insignificant results are the flat budget in P2 and budgets in P4 and P5.

The average winning probability is 0.647 for quack choices and is 0.704 for expert choices, and they are significantly different from each other. This difference is also economically significant under my judgment. In my experiment, the prior probability to win the prize is already high (0.626), and thus it is not easy to improve the winning chance through test choices.

How much subjects can benefit in winning the prize if choosing optimally? Table 3.4 lists the summary statistics of the relative improvement. It is the percentage increment in winning probabilities of the most useful expert test compared to that of the actual chosen test. Since the majority of the expert choices are already on the correct border, there is not much improvement for them. However, subjects who choose quack tests will increase the winning chances by 11.6% on average.

	mean	sd	pt5	pt25	pt50	pt75	pt95
Pool	5.6%	0.074	0%	0%	3.3%	8.3%	21.5%
Quack	11.6%	0.077	3.3%	6.7%	8.3%	16.7%	24.0%
Expert	2.3%	0.047	0%	0%	0%	2.5%	12.7%

Table 3.4: Relative improvements in winning probabilities if choosing optimally

3.5 Unexplained mechanisms

In this section, I discuss the mechanisms behind quack choices. After choosing a test, each participant in my experiment also provides a likelihood estimate of the true state and chooses a bet for each signal realization. This auxiliary task measures how the participant evaluates the test she chose before, and thus provides diagnoses for different channels contributing to quack choices.

3.5.1 Belief updating bias

I first investigate whether subjects are Bayesian and whether their updating biases cause quack choices. In the experiment, subjects will earn a step-wise bonus if their reported posterior of the chosen test is close to the Bayesian posterior. Table 3.5 provides the category of subjects' posterior estimates based on their bonus scales. Subjects are generally Bayesian. 79% of their reported posteriors for both red and white signal are five percent away from the Bayesian ones and thus are entitled to earn a bonus of £1.5. Another 14% of posterior estimates deviate from the Bayesian posteriors by at most fifteen percent and earn a bonus of £0.5. The incentive scheme works equally well for red and white signal.[18]

		Small deviates			Medium deviates	Large deviates
		Bayes	over	under		
Aggregate		26%	24%	29%	14%	7%
Quack	red	76	69	80	34	27
	white	126	62	40	42	16
	pool	35%	23%	21%	13%	8%
Expert	red	97	132	177	78	42
	white	122	133	176	73	22
	pool	21%	25%	34%	14%	6%

Table 3.5: Classification of subjects' posterior estimates. All posteriors are on state L. Category of small deviates include observations that absolute differences between the reported posteriors and the Bayesian posteriors are no larger than five percent. Observations in the category of medium deviates have absolute differences no larger than fifteen percent. Category of "under" means subjects under-estimate state L, and posterior observations over-estimate state L is classified into the category of "over".

The belief updating deviates cannot explain quack and expert choices. Table 3.5 also displays the interaction between updating categories and test choices for each signal. When the signal is red, the posterior deviates of quack and expert tests are not significantly different. But they are significantly different after the white signal.[19] The composition of the category of small deviates indicates that subjects are more likely to under-estimate state L under expert tests. However, it is mainly driven by how experts and quacks are defined, not by subjects' updating abilities.

Figure 3.5 illustrates the distinction between posteriors after the red and white signal. Each panel plots subjects' posterior estimates of state L on the y-axis against the Bayesian posteriors on the x-axis. The dashed lines are from the linear regressions, and the solid lines are from polynomial LOESS (Locally Estimated Scatterplot Smoothing) regressions. Both of them are very close to the identity lines, indicating that subjects are inferring signals as a Bayesian. When the signal is white, Bayesian posteriors for quacks and experts cluster on the separate side of one half. This cluster is consistent with the definition of quacks and experts — a test is a quack if it induces a posterior supports a

[18] The distribution of posterior deviates after red signal is not significantly different from the one after white signal, under both the Wilcoxon signed-rank test and the t-test.

[19] The p-value is 0.0095 under the Wilcoxon rank-sum test and is 0.0145 under the t-test. The result is also valid for absolute deviations.

bet on state L and is an expert otherwise. The lower magnitudes of posteriors for expert tests may contribute to the under-estimation result.

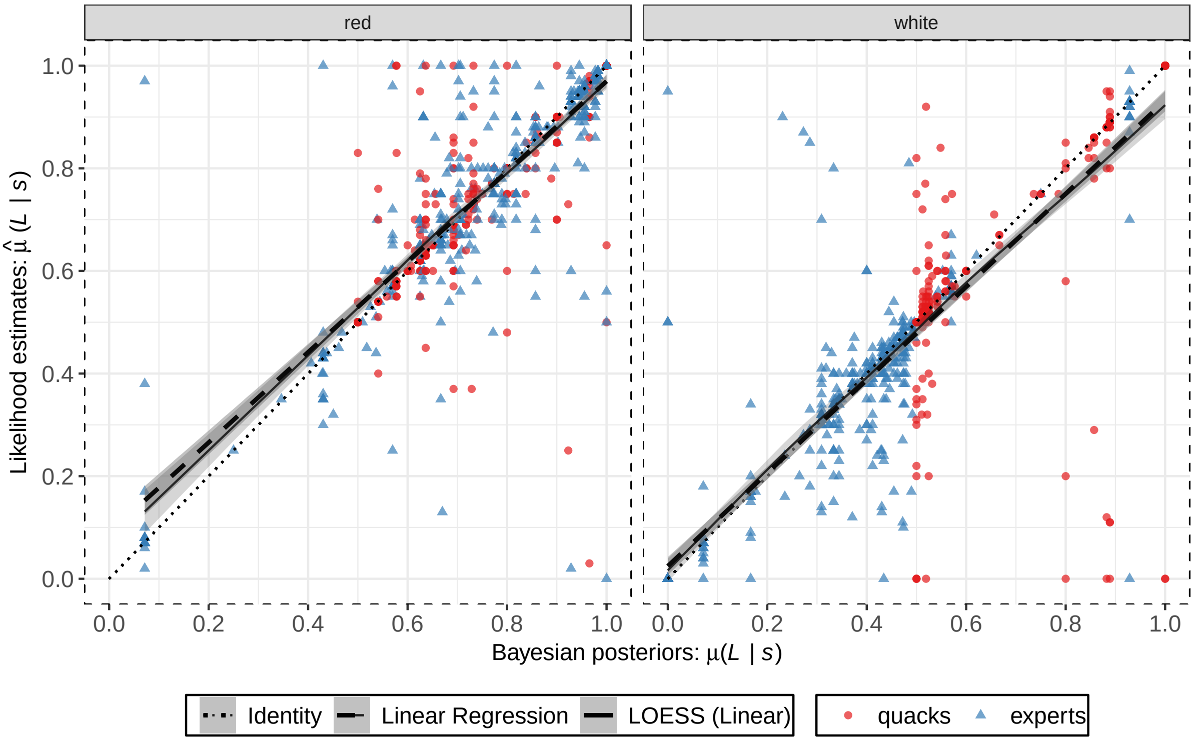


Figure 3.5: Subjects' reported posteriors versus Bayesian posteriors. Shaded regions are 95% confidence intervals for Linear and LOESS regression.

A formal estimation further confirms these findings. Columns (1) and (2) in Table 3.6 report the estimated coefficients and robust standard errors, separately for each signal. The dependent variable is the reported posteriors, and the independent variables are the Bayesian posteriors and a dummy variable for the quack test. The slope for the Bayes posterior is 0.90 and 0.91 for red and white signal, respectively. Both are very close to one, indicating that subjects are generally Bayesian.[20] The coefficients for the interaction terms are not significant for both signals, implying that Bayesian updating bias does not result in quack choices. Since the updating process may also cause the test choices, I use the binary variable of budget types as an instrument for the quack dummy.[21] The estimation results are in columns (3) and (4). Again, their slope coefficients are close to one, and interaction terms are not significant.

[20] I cannot reject the null hypotheses that the slope is equal to one under the χ^2-test. The p value is 0.005 for red signal and is 0.004 for white signal.

[21] The first stage Probit regression shows that a steep budget is highly predictive for quack choices. Having a steep budget, versus a flat one, increases the probability of quack choices by 0.470, with a standard error of 0.065.

To further investigate how the prior and the chosen test affect posterior beliefs separately for quack and expert choices, I update Grether (1980)'s regression and estimate the following model:

$$\ln \frac{\mu_{ij}(L \mid s)}{\mu_{ij}(R \mid s)} = \beta_0 \times D_q + \beta_1 \ln \frac{\mu_j}{1-\mu_j} \times D_q + \beta_2 \ln \frac{\mathbb{P}_{ij}(s \mid L)}{\mathbb{P}_{ij}(s \mid R)} \times D_q + \varepsilon_{ij}.$$

The dependent variable is subject i's posterior odds after signal s in decision problem j. The independent variables include the prior odds for the same problem j, the likelihood ratio under i's chosen test, and a dummy equal to one if the chosen test is a quack. The interaction terms capture whether quacks and experts induce the same extend of specific bias regarding priors and tests. When the subject is Bayesian, the coefficients for both prior odds and likelihood ratio are equal to one. Based on these comparisons, I consider four updating biases (see Benjamin (2019)). The coefficient of prior odds determines whether the subject has a bias of base-rate neglect ($\hat{\beta}_1 < 1$) or confirmation bias ($\hat{\beta}_1 > 1$). The coefficient of likelihood ratio determines whether the subject has under-inference ($\hat{\beta}_2 < 1$) or over-inference ($\hat{\beta}_2 > 1$) bias.

	Dependent: subjective posterior				*Dependent:* ln(subjective posterior odds)									
	OLS		IV		OLS		IV		Split sample: OLS				Split sample: IV	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	P1	(9)	P1	(10)	(11)	(12)
Constant	0.08*	0.02*	0.09**	0.02	0.19	-0.09	0.33*	-0.44**	0.16	0.18	0.20	-0.74**	0.42*	-0.75**
	(0.03)	(0.01)	(0.03)	(0.01)	(0.16)	(0.13)	(0.14)	(0.11)	(0.14)	(0.15)	(0.12)	(0.13)	(0.17)	(0.16)
quack	0.06	0.03			0.14	0.33**				0.33·		1.10**		
	(0.04)	(0.03)			(0.13)	(0.12)				(0.18)		(0.28)		
steep(IV)			-0.01	0.01			-0.18	0.71**					-0.16	0.92**
			(0.04)	(0.03)			(0.14)	(0.13)					(0.19)	(0.18)
Bayes posterior	0.90**	0.91**	0.88**	0.95**										
	(0.04)	(0.03)	(0.04)	(0.03)										
Bayes posterior× quack	-0.08	-0.05												
	(0.06)	(0.06)												
Bayes posterior× steep(IV)			0.005	-0.05										
			(0.05)	(0.07)										
ln(prior odds)					0.99**	1.01**	0.81**	0.59**		0.99**		0.40*	0.70**	0.26
					(0.15)	(0.14)	(0.17)	(0.16)		(0.16)		(0.18)	(0.23)	(0.21)
ln(prior odds)× quack					-0.24	0.01				-0.47·		0.78**		
					(0.24)	(0.21)				(0.27)		(0.25)		
ln(prior odds)× steep(IV)							0.14	0.63**					0.11	0.80**
							(0.22)	(0.22)					(0.29)	(0.27)
ln(likelihood ratio)					0.91**	1.39**	0.90**	1.46**	0.86**		0.96**			
					(0.04)	(0.04)	(0.03)	(0.04)	(0.12)		(0.06)			
ln(likelihood ratio)×quack					-0.04	-0.25**								
					(0.05)	(0.06)								
ln(likelihood ratio)× steep(IV)							0.01	-0.37**						
							(0.04)	(0.07)						
$\ln(\widehat{\text{likelihood ratio}})$										1.07**		1.57**	1.05**	1.55**
										(0.04)		(0.04)	(0.04)	(0.04)
$\ln(\widehat{\text{likelihood ratio}})\times$ quack										-0.05		-0.37**		
										(0.06)		(0.07)		
$\ln(\widehat{\text{likelihood ratio}})\times$ steep(IV)													0.02	-0.40**
													(0.05)	(0.07)
Observations	812	812	812	812	812	812	812	812	116	696	116	696	696	696
Adjusted R^{2}	0.71	0.77	0.71	0.78	0.75	0.83	0.75	0.83	0.45	0.75	0.64	0.84	0.74	0.84
Note:														·p<0.1; *p<0.05; **p<0.01

Table 3.6: Regression results for Bayesian updating bias. Robust standard errors (in parentheses) are clustered at the individual level. Odd columns are estimated with observations after red signal, and even columns are for white signal.

Table 3.6 reports the estimate results in columns (5)–(8), two for OLS, and the other two for IV regression method. When the signal is red, the estimated coefficients of prior odds and likelihood ratio are close to one.[22] All interaction terms are not significant. Subjects in my study slightly exhibit a bias of base-rate neglect and under-inference, but these belief updating biases are not the drive for expert and quack choices. I also estimate these coefficients for the reported posteriors after the white signal. Subjects on average are being Bayesian when the test is an expert, but demonstrate the bias of over-inference when the test is a quack. Instead of having different updating abilities for red and white signal, the distinctive bias patterns are more likely to ascribe to the reverse causality between posteriors after white signal and how quacks are defined. To better identify the effect of priors and tests, I also split the sample and estimated two coefficients separately.[23] The results are in columns (9)–(12). All results are still valid.

The coefficients of Grether regressions provide a convenient illustration of subjects' probability weighting functions in the shape specified in Gonzalez and Wu (1999). The curvature parameter is the same as the coefficient $\hat{\beta}_2$, and the elevation parameter is $(\mu/(1-\mu))^{\hat{\beta}_1-\hat{\beta}_2}$.[24] The paramount evidence of base-rate neglect and under-inference bias in literature often suggests the relationship between the stated posteriors and the Bayesian ones is an inverse S-shaped curve. This is documented in Enke and Graeber (2019). Figure 3.9 in Appendix 3.C.1 depicts the estimated probability weighting curves in my experiment. The curvature and elevation parameters are calculated based on coefficients in columns (5) and (9) of Table 3.6. Since my subjects are generally Bayesian, their stated posterior beliefs for both expert and quack tests are linear with respect to Bayesian posteriors. When the prior is large, they are slightly under-weighting posterior probabilities for quack tests.

[22] These estimates are much larger than the ones in the literature. Benjamin (2019) meta-analyzed results of belief updating biases based on posteriors elicited from Grether (1980)'s bookbag-and-poker-chip experiments. The estimated coefficient $\hat{\beta}_1$ ranges from 0.43 (for all fourteen papers) to 0.61 (for sub-sample of six papers with incentived experiments), and the coefficient $\hat{\beta}_2$ ranges from 0.20 to 0.38. Enke and Graeber (2019) ran a similar belief updating experiment and found the coefficient is 0.52 for prior odds and 0.44 for likelihood ratio. Ambuehl and Li (2018) estimated $\hat{\beta}_2$ at the individual level and found the mean is 0.91 for 143 subjects. I also run Grether regressions for each one of 58 subjects in my experiment. The mean is 0.91 for $\hat{\beta}_1$ and 0.83 for $\hat{\beta}_2$. (see Appendix 3.C.1 for details.)

[23] I first run Grether regressions for the reported posteriors of decision problems in P1. Their prior is one half, making the term regarding prior odds vanishes and providing an estimate of β_2. Then I estimate β_1 based on answers for the rest decision problems in P2-P7. The term replacing likelihood ratio is the fitted value from the first stage regression.

[24] The probability weighting curve of belief posteriors can be very different from the ones used in evaluating lotteries as in Tversky and Kahneman (1979), Tversky and Kahneman (1992), Tversky and Wakker (1995), and Prelec (1998). The ones for lotteries often describe how people perceive the explicitly given probabilities. The one I considered for posteriors incorporates both the perception and belief updating processes.

I also run these regressions for each subject. The results are in Appendix 3.C.1. Table 3.12 provides a summary of the estimation results. Moreover, I divide subjects into two groups, one has quack choices above the median frequency, and the other half has more expert choices. Figure 3.10 compares their distributions of each estimated coefficient. The group with more expert choices is slightly closer to Bayesian updating than the other group. In general, subjects are Bayesian, and two groups of subjects do not exhibit much difference in belief updating biases.

3.5.2 Sub-optimal bet choices

Given their posteriors of the state, are subjects choosing the optimal bet? The failure of best-responding to their own beliefs may also contribute to quack choices. Costa-Gomes and Weizsäcker (2008) provide evidence for such failures in strategic games. Subjects in their experiment play two-person games and state their beliefs about what their opponents will do. They find players' strategy is not consistent with their stated beliefs for more than half of the games. In my experiment, subjects face individual decision-making problems. Their beliefs and bets are both on the true state of the world, and it is straightforward for them to infer optimal bets from beliefs. Indeed, subjects' bet choices in auxiliary tasks are best responding to their reported posteriors.

Table 3.7 reports the distribution of subjects' bet choices, clustered for expert and quack tests and different signals. When the signal is red, subjects recognize that the optimal bet is on state L regardless of the chosen test. The rate is 95% under quack and 92% under expert test. The bet choices are largely consistent with their reported beliefs and the corresponding Bayesian beliefs. When the signal is white, subjects predominantly bet state L under quacks (71%) and state R under experts (92%). This finding confirms the failure in distinguishing experts and quacks. It also suggests that people did not realize that a test is useless if it leads to the same action for both signals. However, subjects' bet choices are still highly consistent with their reported beliefs and the Bayesian posteriors, implying the bias in best-responding does not drive quack choices. Table 3.13 in Appendix 3.C.2 further reports the number of belief-inconsistent bet choices for both quack and expert tests. Overall, only 3% of bets are strictly inconsistent with the reported and the Bayesian posteriors, and those made under quack and expert tests each take around 1.5%.

I also consider two alternative definitions of experts and quacks. One is based on subjects' reported posteriors, and the other is on subjects' bet choices. A test is a "belief quack" (or "bet quack") if the expected winning probability of having the test is the same as the prior according to the DM's reported beliefs (or bet choices).

			Bet L	Indifferent	Bet R
Quack	red	empirical	271	—	15
		under reported belief	256	23	7
		under Bayesian belief	262	24	0
	white	empirical	202	—	84
		under reported belief	178	77	31
		under Bayesian belief	203	83	0
Expert	red	empirical	483	—	43
		under reported belief	473	3	50
		under Bayesian belief	482	0	44
	white	empirical	42	—	484
		under reported belief	54	8	464
		under Bayesian belief	44	0	482

Table 3.7: Distributions of empirical bet choices and those consistent with the reported and Bayesian posterior beliefs

Compared to the rational DM who updates beliefs accurately and bets optimally, each definition relaxes an aspect of rationality with subjects' own evaluations of the chosen test. The belief-based definition admits subjective inference in the process of evaluating tests, and the action-based definition focuses on the consequences of choosing tests. Figure 3.6 provides the histogram of quacks choices under each definition. In summary, frequencies of quacks are robust under various definitions, confirming that neither the belief updating bias nor the best-responding bias explains quack choices.

3.5.3 Intrinsic preferences over test characteristics

I continue to examine whether intrinsic preferences play a role in quack choices. If subjects care about a specific attribute, and tests on quack segments are more likely to have the attribute than those on expert segments, many test choices will be quacks. For example, if a DM has a strong preference for symmetric accuracies, all of her test choices are quacks since the default symmetric tests always induce the same winning probability as the prior. In the literature, intrinsic preferences often capture a taste for skewness. Masatlioglu, Orhun and Raymond (2017) find people prefer positively skewed test[25] in an experiment with non-instrumental information. Their definitions of skewness are not applicable in my setting. With the instrumentally valuable information, subjects

[25] In their setting, a test is positively skewed if it resolves more uncertainty about the high outcome than about the low outcome.

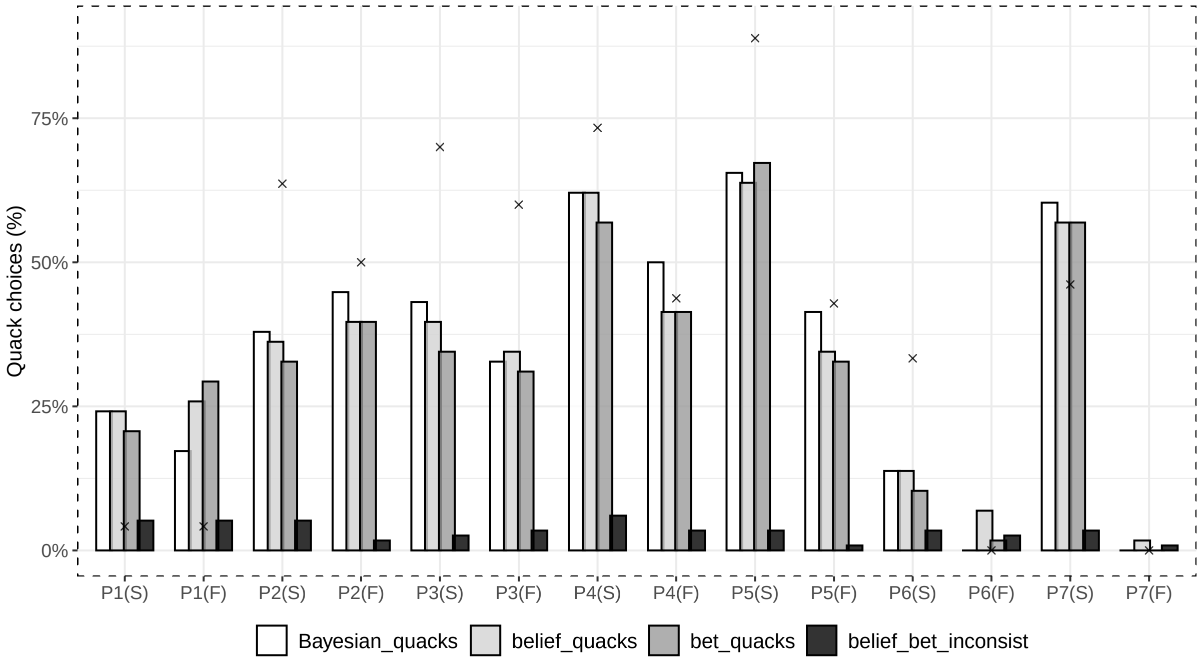


Figure 3.6: The histogram of quack choices under alternative definitions.

choose bet for each possible signal realizations, and such action contingency blends the desired and undesired outcomes. In other words, there is no high or low outcome in my experiment because both bets on state L and R have chances to win a same prize. Moreover, I elicit preferences over tests through pairs of steep and flat budgets. Even though all priors favor state L, these budget pairs produce similar opportunities to resolve state L and R.

I consider asymmetry measures based on the test (p, q) itself and its induced posterior distribution (μ_r, μ_l). The test-specific measures include the absolute difference between two accuracy levels $|p - q|$ and its distance to the default symmetric test $\|(p, q) - \text{pivot}\|$. The posterior-specific asymmetry is described by the unconditional probability of red signal. This is due to the martingale property of belief updating $\mu_l \mathbb{P}(\text{red}) + \mu_r (1 - \mathbb{P}(\text{red})) = \mu$. Since people also care about how asymmetric a test is relative to alternative ones, I consider a relative version for each asymmetry measure: the accuracy ratio q/p, the slope $(q - \text{pivot})/(\text{pivot} - p)$, and the relative belief updating ratio $(\mu_l - \mu)/(\mu - \mu_r)$.[26]

[26] Bordalo, Gennaioli and Shleifer (2012)'s salience theory provides explanations for the preference over both absolute and relative skewness for risk lotteries. Studies on prudence (see Trautmann and van de Kuilen (2018) for a survey) show that people prefer positively skewed lotteries. Dertwinkel-Kalt and Köster (2019) provides experimental evidence for the preference over relative skewness for choices under risk.

Table 3.8 summarizes six asymmetry measures. They vary a lot. Some of them like the accuracy ratio and the relative belief updating ratio are close to the symmetric benchmark of one. Other measures document subjects' preferences over asymmetries. For instance, subjects choose color compositions generating slightly more red balls than white balls for Box A. Yet, most measures are comparable between expert and quack tests. Table 3.14 in Appendix 3.C.2 reports Probit regression results of test choices on each of the asymmetry measure, and none of them predict quack choices. Taken together, I find people exhibit similar preferences for variously defined asymmetry measures. Therefore, intrinsic preferences over test characteristics do not justify quack choices.

		mean	sd	pt5	pt25	pt50	pt75	pt95
$\lvert p-q \rvert$	quack	0.51	0.33	0.00	0.23	0.57	0.82	0.88
	expert	0.47	0.28	0.00	0.23	0.42	0.75	0.90
$\lVert ((p,q)-\text{pivot}) \rVert$	quack	0.41	0.28	0.00	0.18	0.45	0.69	0.76
	expert	0.36	0.20	0.00	0.20	0.34	0.53	0.67
$\mathbb{P}(\text{red})$	quack	0.65	0.24	0.17	0.49	0.64	0.89	0.92
	expert	0.62	0.25	0.24	0.43	0.55	0.90	0.97
q/p	quack	0.94	1.13	0.00	0.06	0.70	1.44	3.57
	expert	1.02	0.72	0.10	0.25	1.12	1.54	2.50
$(q-\text{pivot})/(\text{pivot}-p)$	quack	3.54	1.93	0.88	1.50	4.00	6.00	6.00
	expert	3.09	1.73	0.88	2.00	2.50	4.00	6.00
$(\mu_l-\mu)/(\mu-\mu_r)$	quack	1.01	1.43	0.09	0.13	0.56	1.05	4.95
	expert	0.94	0.87	0.03	0.11	0.82	1.31	3.17

Table 3.8: Summary of asymmetry measures of quack and expert tests. The top six rows are based on absolute measures of asymmetry, and the bottom six rows are on relative measures.

3.6 Explained mechanisms: contingent reasoning and decision rules

This section investigates the remaining channel for quack choices: the failure of contingent reasoning. It is an inference bias occurred when evaluating the usefulness of a test, and it is unobservable. I study people's reasoning bias based on their descriptions on how they chose color compositions.

Table 3.15 in Appendix 3.C.2 lists subjects' comments. They describe how subjects reason the color and bet choices in the experiment.[27] Three decision rules are popular among subjects. The first one is to choose tests whose accuracies are either 0% or 100%, therefore, subjects can resolve the uncertainty about the state for one of the signals. The second rule is to make the coloring of box L and R as different as possible. This rule is intuitive to subjects when they move sliders and observe the dynamic changes of two boxes. The third one is to maximize the chance of one signal so that subjects are more certain about which one of two bets counts.[28] I refer to these reasoning processes as *entropy-reducing*, *evidence-separating*, and *signal-separating* rules.

All three decision rules favor tests on the borders of (p, q) space. I illustrate this point through coloring representations of border tests in Table 3.9. Each entry plots a simplified version of how box L and R and the induced posteriors look like under the considered border test. Tests on the top-left border allocate as many as white balls for box R until it is full, bringing about a certainty of state L after a red signal, a noticeable difference between two boxes in their coloring, and a highest chance to receive a white signal. These decision rules also justify the choices of tests on the bottom-right border. However, I construct the decision problems with paired budgets such that none of the decision rules guarantees expert (or quack) choices. Both types of border tests can be experts or quacks, depending on the budget. When it is steep, accuracy q is relatively cheaper than p, making the top test $(1/2, 1)$ the most useful expert and the bottom test $(3/4, 0)$ a quack. The opposite holds for flat budgets. The key to distinguish the expert and the quack between two border tests still hinges on whether the test induces a pooling or separating bet choices. In fact, all decision rules acknowledge the optimal expert tests if subjects reason contingently.

My classification of decision rules are different from ones considered in the literature due to the unique approach to construct the choice set of tests. For example, in Charness, Oprea and Yuksel (2018) and Montanari and Nunnari (2019)'s experiment, subjects choose one from two test options, one biases towards the state favored by the prior, and the other one biases against it. Charness, Oprea and Yuksel (2018) consider the symmetric case where the prior-confirming test is $(1, \lambda)$, and the prior-contradicting

[27] I did not report uninformative comments or those do not describe decision processes such as "I calculated probabilities" and "I made a guess".

[28] Typical comments for these decision rules are: (1)"I made sure that wherever I could, there was an option that red or white would 100% be label R or L" from the subject with ID 59; (2) "The colour choices are based on the difference in red and white between L and R, you make the gap as big as possible so its easier to choose L or R from red and white. The bet is then based on the ratio between the gaps." from ID 8; and (3) "Try to favor one colour, increasing the chances for one colour to have a high change to belong to one of the boxes" from ID 43.

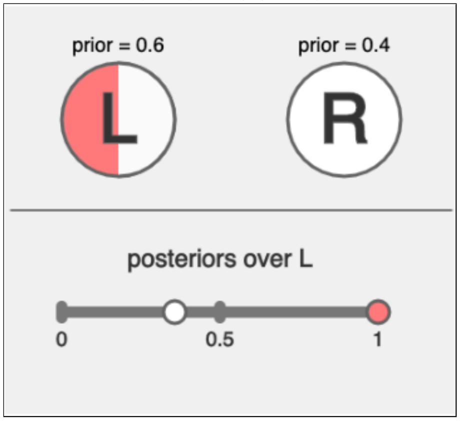
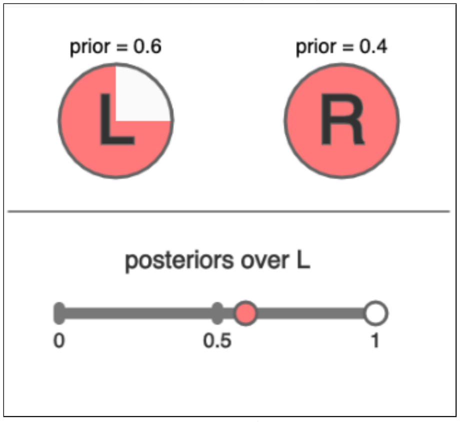
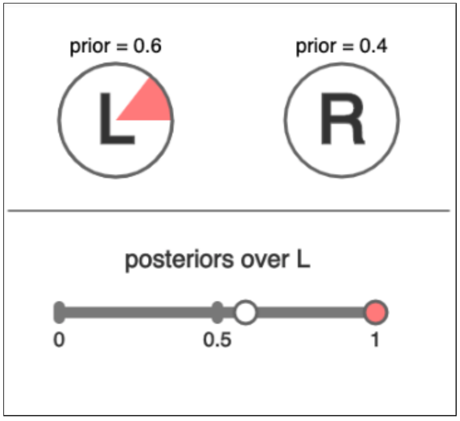
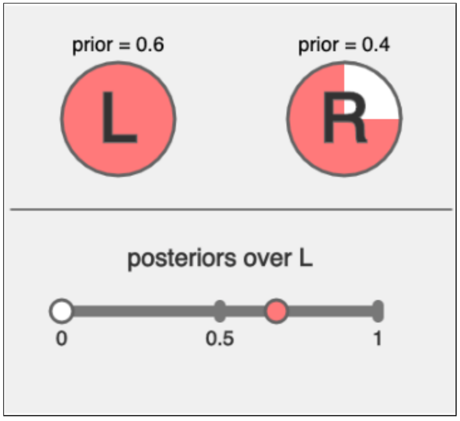
	Top-left border	Bottom-right border
Steep budget	$(p,q) = (1/2,1)$	$(p,q) = (3/4,0)$
q cheaper	 prior = 0.6 prior = 0.4 L R posteriors over L 0 0.5 1	 prior = 0.6 prior = 0.4 L R posteriors over L 0 0.5 1
	expert	quack
Flat budget	$(p,q) = (1/7,1)$	$(p,q) = (1,1/4)$
p cheaper	 prior = 0.6 prior = 0.4 L R posteriors over L 0 0.5 1	 prior = 0.6 prior = 0.4 L R posteriors over L 0 0.5 1
	quack	expert

Table 3.9: An illustration of border tests and their induced posterior distributions for the steep and flat budget in P3. The expert test on the top-left border of the steep budget shares the same value as the one on the bottom-right border of the flat budget.

test is $(\lambda, 1)$. Montanari and Nunnari (2019) consider the asymmetric case where one is $(1, \lambda)$, and the other one is $(1-\lambda, 1)$. Both cases coincide with two border tests of the flat budget in my setting. The top-left test is prior-contradicting, and the bottom-right one is prior-confirming. In addition, each one of them corresponds to a comparable border test on the steep budget. For instance, the accuracies of two expert tests are carefully designed to generate the same extent of usefulness for betting decisions. Such constructions allow for a new identification for the relationship between test choices and decision rules.

Since all decision rules rationalize border tests, I quantify each one of them by an attribute of border tests. Each considered attribute reflects how the corresponding rule favors border tests. First, I measure the decision rule that aims to reduce overall uncertainty by the decrease of entropy from the prior to the induced posterior. It is based on Shannon (1948)'s entropy concept for the uncertainty inherent in a random variable, and further developed in Cabrales, Gossner and Serrano (2013) for the uncertainty induced by an information structure.[29] I measure the evidence-separating rule by the coloring differences between two boxes. More specifically, it equals to $|p-(1-q)|$, the absolute gap between the proportion of red (or white) balls for each box given test (p,q). The signal-separating rule is calibrated by the unconditional probability of the dominating signal, which is $\mathbb{P}$(red) for top-left tests and $\mathbb{P}$(white) for bottom-right tests.

I examine how these rule-specific measures affect subjects' test choices through reduced-form Probit regressions. Table 3.10 reports the results. Columns (1)–(3) show that measures for each decision rule are highly predictive for expert versus quack choices.[30] The chosen test is more likely to be a quack if the top test on the same budget resolves more uncertainties, or both the top and bottom tests generate larger differences in the coloring between two boxes. A bottom test allocating more white balls to Box A predicts a higher chance of choosing an expert test. It is very likely that three decision rules influence quack (or expert) choices via the indirect channel of influencing choices on the top or bottom segments of budgets (see regression in column (4)). I regress a dummy variable for tests on the top segments over the measures for each rule in columns (5)–(7). Subjects' choices on the top segment are significantly affected by entropy-reducing and evidence-separating rules, but not signal-separating rule.[31]

I predict the rates of choosing quacks if subjects use entropy-reducing or evidence-separating rules to choose choices on top segments. This is possible due to my paired construction of budgets under which all choices on top segments are experts when budgets are steep, and they are quacks when budgets are flat. Figure 3.7 compares the actual quack rates with the predicted ones across budgets in P2-P7. Both decision rules

[29] Here is a formal calculation of entropy reduction. Denote the Shannon entropy of an arbitrary belief distribution as $H(\mu') = -(\mu' \log_2(\mu') + (1-\mu')\log_2(1-\mu'))$, where μ' is belief (prior or posterior) over state L. For a test (p,q), its entropy is an expectation of $H(\mu_s(p,q;\mu))$, weighted by the unconditional probability of each signal. The expected decrease of entropy from the prior to the posterior induced by (p,q) is $\Delta H(p,q;\mu) = H(\mu) - \sum_s H(\mu_s(p,q;\mu))\mathbb{P}(s)$.

[30] The coefficients of top and bottom attributes are jointly significant in each one of the regressions under χ^2-test. p-values are 0.044 for entropy-reducing measures, 0.000 for both evidence-separating and signal-separating measures.

[31] Joint hypotheses tests on the coefficients of top and bottom attributes show that p-values are 0.032 and 0.015 respectively for the entropy-reducing and evidence-separating rule. p-value is 0.619 for attributes measuring the signal-separating rule.

	Dependent: D(expert choice)				*Dependent:* D(top choice)		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Constant	-1.83	-45.74**	-3.46	0.13*	-5.69*	-22.66**	0.56
	(2.36)	(8.43)	(6.13)	(0.06)	(2.35)	(7.82)	(5.17)
Slope	0.84	10.60**	-1.12		1.40**	5.44**	-0.11
	(0.52)	(1.91)	(1.15)		(0.51)	(1.75)	(0.97)
Size	-1.33$\cdot$	-14.88**	0.97		-2.27**	-7.81**	-0.19
	(0.72)	(2.66)	(1.51)		(0.71)	(2.44)	(1.27)
Quack chance	-3.72**	-3.52**	-2.72**		-1.59**	-1.00**	-1.18*
	(0.52)	(0.38)	(0.69)		(0.48)	(0.30)	(0.58)
Steep	0.89*	2.48**	0.85**		1.34**	1.71**	1.01**
	(0.41)	(0.47)	(0.29)		(0.41)	(0.45)	(0.29)
Pivot point	9.32$\cdot$	104.96**	7.74		13.53**	50.36**	-0.98
	(4.99)	(18.57)	(11.79)		(4.88)	(17.05)	(9.85)
D(Top choice)				0.44**			
				(0.10)			
Top: Δ(entropy)	-5.28*				-4.50*		
	(2.22)				(2.10)		
Bottom: Δ(entropy)	-3.44				-4.38*		
	(2.19)				(2.17)		
Top: $\lvert p+q-1 \rvert$		-25.89**				-11.03**	
		(4.54)				(4.10)	
Bottom: $\lvert p+q-1 \rvert$		-13.58**				-7.42**	
		(2.87)				(2.60)	
Top: $\mathbb{P}$(red)			-2.30				2.03
			(3.88)				(3.20)
Bottom: $\mathbb{P}$(white)			12.55**				2.27
			(2.89)				(2.66)
Observations	696	696	696	696	696	696	696
Note:					$\cdot$p<0.1; *p<0.05; **p<0.01		

Table 3.10: Probit regression results for three heuristics

predict distributions of quacks similar to the actual one. On average, the predicted quack rates are slightly higher, but none of them are significantly different from the actual quack rate under *t*-test and Wilcoxon rank sum test. This finding implies that quack choices are simply by-products of choices under simple decision rules. Furthermore, it confirms the universal failures of contingent reasoning. When facing the interaction between test choices and bet choices, people fail to reason the distinction between useful and useless tests, and they use simple decision rules aiming to reduce overall uncertainties or differentiate evidence structures.

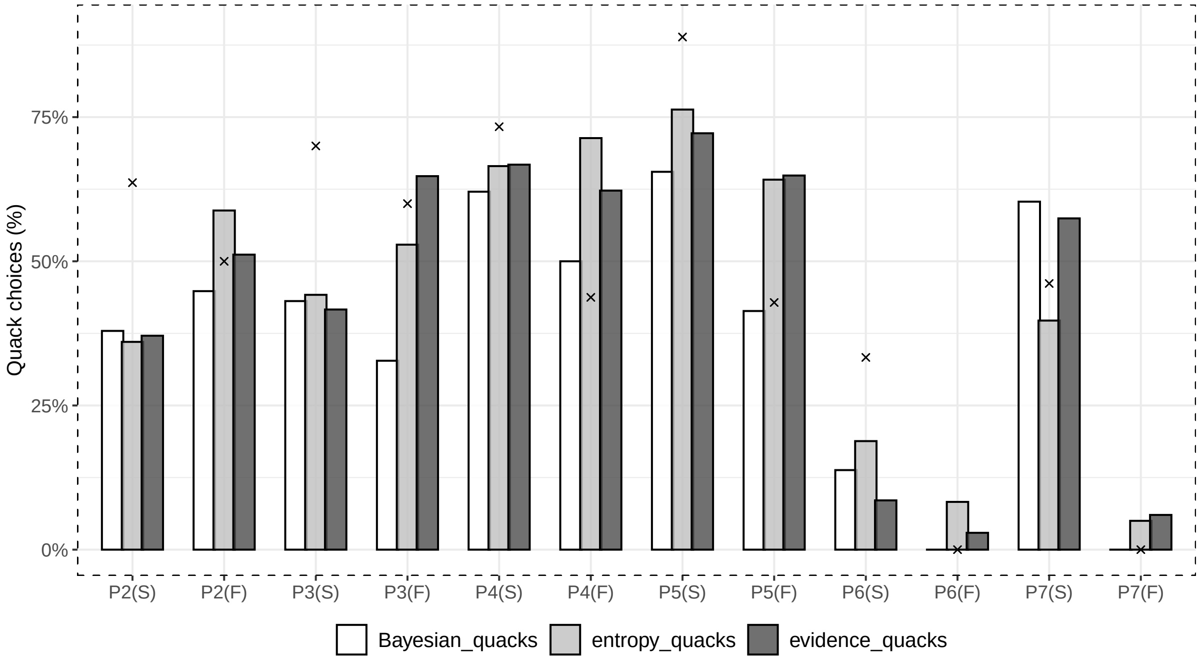


Figure 3.7: The histogram of predicted quack choice rate for budgets in P2-P7.

Notice that theses decision rules are not substitutes to contingent reasoning. It is not the case that subjects turn to simple decision rules because they find it difficult to reason contingently. For example, under the entropy-reducing rule, a DM cares about how much the induced posterior distributions resolve the overall uncertainties about the state. However, such consideration does not internalize the contingent reasoning in how expert and quack tests affect the optimal actions in different ways. Similarly, neither evidence-separating nor signal-separating rule considers the influence of tests on actions. In other words, contingent reasoning is not exclusive to decision rules. A subject who reasons contingently and employs each one of the rules will choose the optimal test. Since these rules favor the salient tests on the border, many subjects either choose the most useful experts or the most distant quacks.

Additional analyses on post-experiment questionnaires show that quack choices are correlated with individual cognitive abilities. Table 3.16 in Appendix 3.C.3 reports regression results of subjects' quack choice rate over demographic variables, self-assessed attitudes, and cognitive scores. Subjects with higher CRT scores are choosing less quacks, while the other two psychometric measures are not significant. Self-reported attitudes include risk aversion and having strong opinions predict more quack choices. Surprisingly, subjects who self-reported to be good at figuring out useful clues are also making more quack choices. The attitudes determining quack choices are not the same as the ones for Bayesian updating. I run similar regressions for individual measures of belief updating bias in Table 3.17. The CRT score is still significant. In addition, being male and the self-reported ability to have different perspectives predict a lower updating bias.[32]

3.7 Discussion and implication

In my setting, individuals choose tests to reduce risks over two payoff relevant states. I characterize the state-outcome correspondence by two bets having symmetric outcomes. Under this construction, individuals' preferences over outcomes are irrelevant to their decisions onn tests. The outcome π can be non-monetary. All subjects compare their beliefs with the fixed threshold of one half. One implication is that my setup is immune to individuals' utility-specific risk attitudes. The threshold has already encapsulated different shapes of utility functions over outcomes.[33] When the prizes for two states are not symmetric, for instance, the case of choosing medical tests, we can elicit each individual's threshold beforehand.[34] All theoretical analyses and the experiment remain valid to study people's choices of tests and reasoning biases.

The contingent reasoning bias is not restricted to a setting with binary signals or binary states.[35] For decision problems with non-binary signals, the DM updates beliefs

[32] In unreported tables, CRT score and perspective ability are still significant when the measures for individual updating biases are coefficients estimated from Grether regressions.

[33] Individuals' probability-specific risk attitudes are often reflected by different shapes of probability weighting functions. When evaluating tests, people's reported posterior probabilities incorporate how they inference signals and how they perceive probabilities. Therefore, people's probability-specific risk attitudes and inference biases are indistinguishable from each other. I generically refer to the deviation from the Bayesian posteriors as belief-updating biases.

[34] A DM's decision threshold is her belief over state L that makes her indifferent between bets on state L and R. We can easily elicit the threshold by matching the probability of a bet with the high prize with the option of receiving the low prize for sure.

[35] Notice all discussions of states and signals admit standard assumptions in the literature. The state space is finite. The action space of bets is the same as the state space. The size of the signal space is not smaller than that of the state space.

for each signal and evaluates tests based on signal-contingent winning probabilities. When there are more than two states, the DM chooses bets according to multiple thresholds, each corresponding to an indifference between two bets. All thresholds partition the DM's belief space into different regions, and beliefs in each region support one particular bet being optimal.[36] None of the extensions fundamentally changes the evaluation process of tests. A test is an expert when it induces posteriors spreading at different belief partitions, and it is a quack if it induces posteriors supporting the same bet as in prior.

The supply side of tests may also submit to the contingent reasoning bias. Many studies in contract theory and mechanism design are related to the optimal disclosure of information from an ex ante perspective. For instance, Lizzeri (1999), Eső and Szentes (2007*a*), Eső and Szentes (2007*b*), and Li and Shi (2017) consider the case where a seller knows customers' match-value of a good and commits to a disclosure rule (before knowing the type of the customer) and a price policy to maximize revenues (or screen customers). Bergemann, Bonatti and Smolin (2018) studies the sales of ex ante information.[37] A data seller owns a database regarding the states and offers data buyers different versions of statistical experiments (tests) before the realization of the true state. A seller who cannot anticipate how different tests influence customers' decisions may not be able to design the optimal contract or information products. On the other hand, a seller may also take advantage of buyers' reasoning bias and obfuscate them with quack tests on purpose. Our findings are relevant to policy regulations on such markets.

The contingent reasoning principle also applies to a strategic setting. A sender chooses a test, and a receiver learns a realization of the signal and then take action. Their interests are not aligned, motivating the sender to choose particular tests to serve his ends. This setting relates to an extensive literature on information design. For example, Myerson (1991) analyzes the value of communication in sender-receiver games and illustrates how a mediator improves information transmission via an appropriately chosen test. A sender in Kamenica and Gentzkow (2011)'s Bayesian persuasion model commits to an information structure to persuade a receiver to choose the action the sender prefers as likely as possible.[38] More discussions about different mechanisms of information design and their connections can be found in Bergemann and Morris (2019).

[36] Formal proofs are out of the scope of this paper. Interested readers are referred to Kamenica (2019)'s discussion of the concavification approach for the value function of information structures and Lara and Gossner (2020)'s convex analyses for the duality between payoffs and posteriors.

[37] Bergemann and Bonatti (2019) provides a recent review on the market for both ex ante and ex post information.

[38] Kamenica (2019) provides a recent survey on Bayesian persuasion models. A review of the empirical work on persuasion can be found in DellaVigna and Gentzkow (2010).

My findings suggest an important yet often neglected reasoning requirement for this strand of literature. The reasoning biases from either the sender or the receiver side have concrete implications for both the theoretical validity and the empirical implementation of these mechanisms.

3.8 Conclusion

This paper provides a unified framework of expert and quack tests and studies how people choose and evaluate them. Using a budget experiment, I document the failure in distinguishing experts and quacks and examine underlying mechanisms. People frequently select quacks even though many decision-enhancing expert tests are also in their choice set. This finding is not explained by belief-updating bias, sub-optimal action choices, or intrinsic preferences over test characteristics. People choose quacks because of the failure of contingent reasoning in information processing.

This reasoning bias has profound consequences for many decision problems. It leads to demand (or supply) for inferior information sources or obfuscating disclosure policies at the beginning of decision-making. Once a DM chooses a quack test, all her efforts in correcting cognitive biases and choice mistakes will be in vain. However, the cost of de-biasing can be meager. As long as the DM realizes the interaction between tests and their influences on decision problems, she will choose optimally even under simple decision rules. This paper also suggests that de-biasing interventions targeted to reasoning processes may be more beneficial than those targeted to belief or action biases. How to educate people to reason contingently is worth future study.

Appendix 3.A Proofs

Proposition 3.1. *An admissible test induces a Bayesian posterior spread such that observing signal l (or r) increases the posterior of the state L (or R) relative to the prior, and vice versa. Mathematically, for any prior μ, a test $(p,q) \in \mathcal{T} \iff \mu_r^{Bayes}(p,q;\mu) \leq \mu \leq \mu_l^{Bayes}(p,q;\mu)$.*

Proof. Given a prior $\mu \in [1/2, 1)$ and an arbitrary test (p,q), the Bayesian posteriors after signal l and r are

$$\mu_l^{Bayes} = \frac{p\mu}{p\mu + (1-\mu)(1-q)} \quad \text{and} \quad \mu_r^{Bayes} = \frac{(1-p)\mu}{(1-p)\mu + q(1-\mu)}.$$

The inequality $\mu_l^{Bayes} \geq \mu$ holds if and only if $p + q \geq 1$, implying the test is admissible. Analogously, the inequality $\mu_r^{Bayes} \leq \mu$ is also equivalent to the admissible condition. □

Proposition 3.2. *Given a rational DM and a prior μ, an admissible test (p,q) is an expert (a quack) if and only if the condition $(1-p)\mu - q(1-\mu) < 0$ is (not) satisfied. The value of an expert test is $v^{Bayes}(p,q;\mu) = p\mu + q(1-\mu)$, and the value of a quack test is μ.*

Proof. In my setting, a test is an expert for a rational DM if and only if the Bayesian posterior belief after signal r is smaller than one half. Solving $\mu_r^{Bayes} < 1/2$ yields the inequality

$$(1-p)\mu - q(1-\mu) < 0.$$

With the inequality $\mu_r^{Bayes} < 1/2$, the rational DM's ex-ante winning probability is

$$\begin{aligned} v^{Bayes}(p,q;\mu) &= \mu_l^{Bayes}\mathbb{P}(s=l) + (1-\mu_r^{Bayes})\mathbb{P}(s=r) \\ &= \mathbb{P}(l \mid L)\mathbb{P}(L) + \mathbb{P}(r \mid R)\mathbb{P}(R) \\ &= p\mu + q(1-\mu). \end{aligned}$$

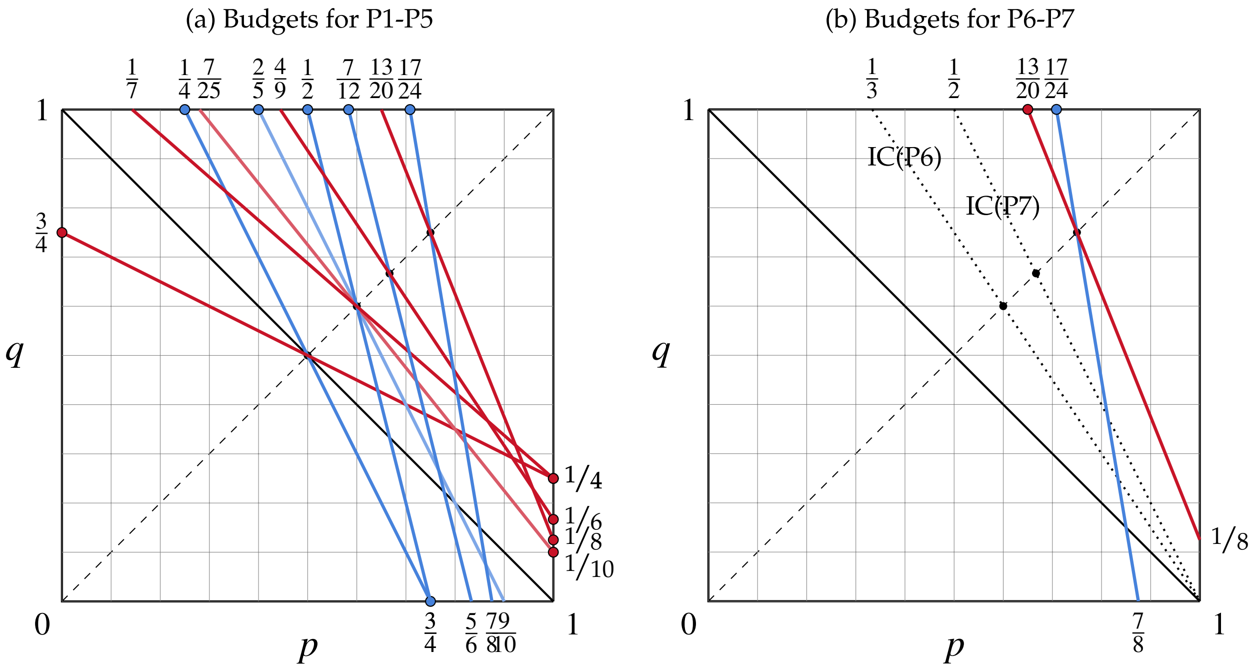
On the contrary, a test is a quack for a rational DM if and only if $\mu_r^{Bayes} \geq 1/2$. With this inequality, the value of the test is

$$\begin{aligned} v^{Bayes}(p,q;\mu) &= \mu_l^{Bayes}\mathbb{P}(s=l) + \mu_r^{Bayes}\mathbb{P}(s=r) \\ &= \mathbb{E}_s \mu_s^{Bayes} \\ &= \mu. \quad \text{(iterated law of expectations)} \end{aligned}$$

□

Appendix 3.B More details on the experimental design

Figure 3.8: Fourteen linear budgets for the experiment. The solid dotes on each budget indicate the optimal choice of the test. Two dotted lines in the right panel are the indifference curves for budget pair P6 and P7 respectively.



(a) Budgets for P1-P5

(b) Budgets for P6-P7

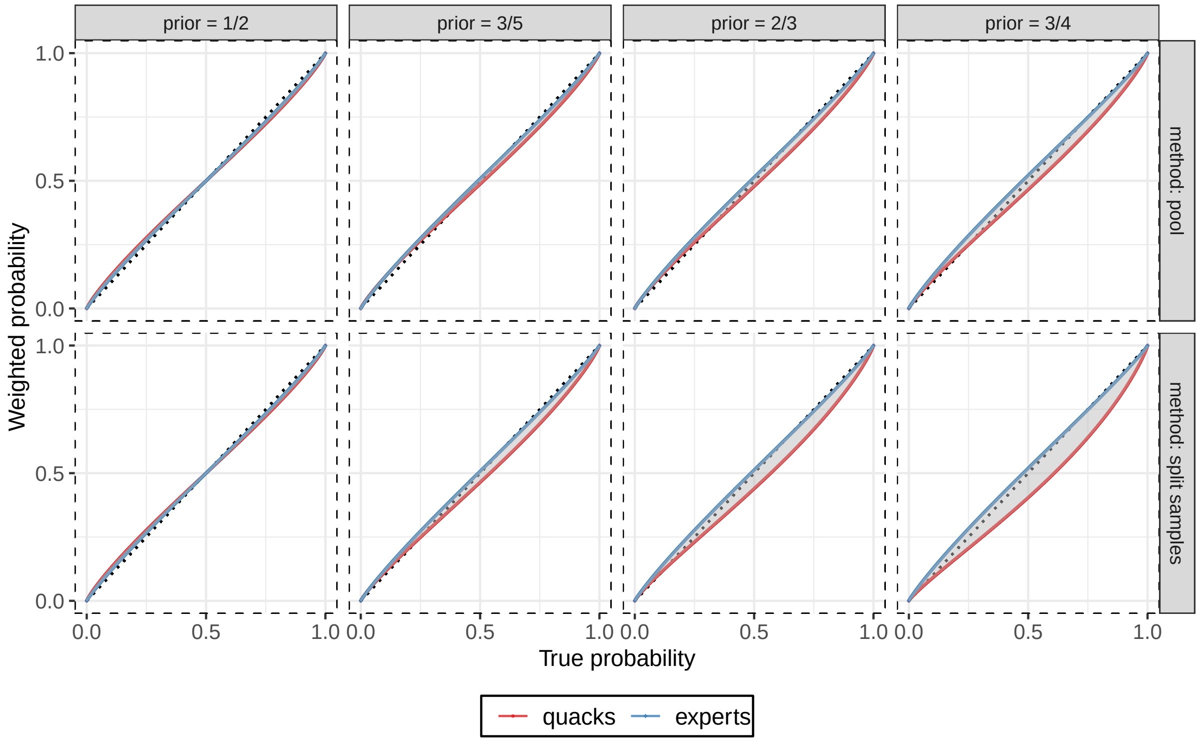
Table 3.11: Constructing Box L and R for each budget. N(L) is the number of balls in Box L. Step(L) is the precision of Box L's slider, indicating the minimum numbers of balls can be colored. All probabilities are discretized based on these natural frequencies.

No.	(Prior, Pivot)	Budget	N(L)	N(R)	Step(L)	Step(R)
P1(S)	$(1/2, 1/2)$	$2p+q=3/2$	50	50	1	2
P1(F)	$(1/2, 1/2)$	$p/2+q=3/4$	50	50	2	1
P2(S)	$(3/5, 3/5)$	$2p+q=9/5$	135	90	3	4
P2(F)	$(3/5, 3/5)$	$5p/4+q=27/20$	150	100	6	5
P3(S)	$(3/5, 3/5)$	$4p+q=3$	135	90	3	8
P3(F)	$(3/5, 3/5)$	$7p/8+q=9/8$	150	100	12	7
P4(S)	$(2/3, 2/3)$	$4p+q=10/3$	120	60	2	4
P4(F)	$(2/3, 2/3)$	$3p/2+q=5/3$	120	60	4	3
P5(S)	$(3/4, 3/4)$	$6p+q=21/4$	120	40	2	4
P5(F)	$(3/4, 3/4)$	$5p/2+q=21/8$	120	40	6	5
P6(S)	$(3/5, 3/4)$	$6p+q=21/4$	60	40	1	4
P6(F)	$(3/5, 3/4)$	$5p/2+q=21/8$	120	80	3	5
P7(S)	$(2/3, 3/4)$	$6p+q=21/4$	60	40	1	3
P7(F)	$(2/3, 3/4)$	$5p/2+q=21/8$	120	60	4	5

Appendix 3.C Additional analyses

3.C.1 Additional analyses on belief updating bias

Figure 3.9: Probability weighting functions estimated from the coefficients in Grether regressions.



3.C.2 Additional analyses on sub-optimal bet choices and test characteristics

Table 3.12: Summary of belief updating regression results at the individual level. The coefficients are for Bayesian posteriors ($\hat{\alpha}_1$) in OLS regression, log prior odds ($\hat{\beta}_1$) and log likelihood ratio ($\hat{\beta}_2$) in Grether regression. The quantile a is denoted as pt$a$.

		mean	sd	pt5	pt25	pt50	pt75	pt95
pool	Bayes posterior	0.94	0.15	0.65	0.91	0.99	1.00	1.06
red	Bayes posterior	0.87	0.24	0.34	0.80	0.98	1.00	1.03
white	Bayes posterior	0.89	0.41	0.31	0.94	1.00	1.00	1.09
pool	ln(prior odds)	0.61	0.34	0.09	0.40	0.58	0.82	1.18
	ln(likelihood ratio)	1.06	0.32	0.46	1.04	1.15	1.21	1.31
red	ln(prior odds)	0.91	0.90	-0.04	0.56	0.93	1.05	2.16
	ln(likelihood ratio)	0.83	0.53	-0.11	0.85	0.99	1.02	1.08
white	ln(prior odds)	0.59	1.24	-1.44	0.39	0.83	1.20	1.69
	ln(likelihood ratio)	1.16	0.68	0.47	1.17	1.33	1.39	1.49

Figure 3.10: The mean and 95% confidence intervals of the estimated coefficients in OLS regression for Bayesian posteriors and Grether regression for likelihood ratio and prior odds. The coefficients are based on subjects' posteriors after the red signal.

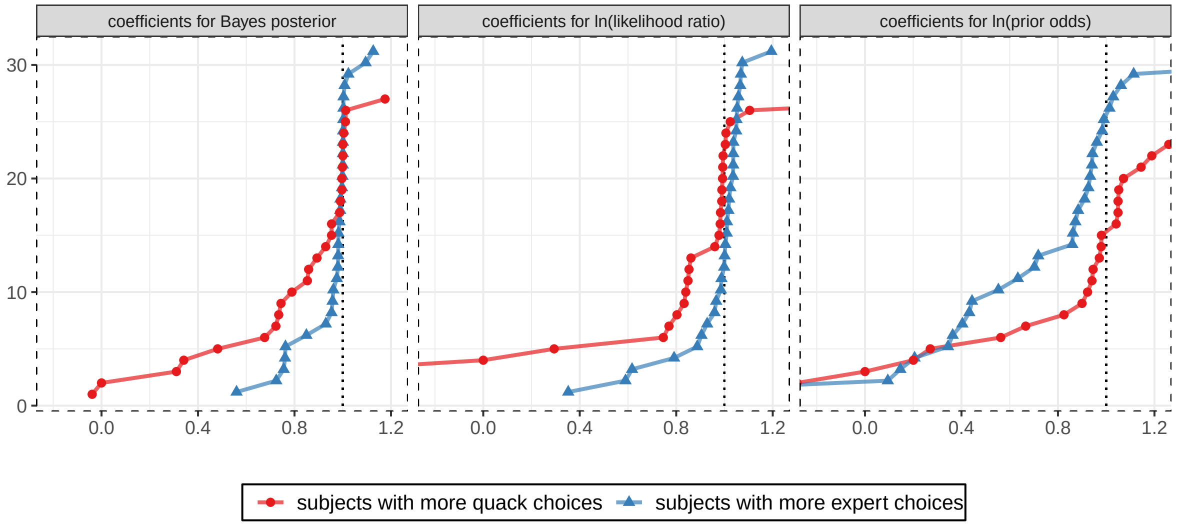


Table 3.13: Counts of bet choices that are strictly inconsistent with the reported beliefs and the Bayesian beliefs

		Under reported belief		Under Bayesian belief	
		quack	expert	quack	expert
Red	incorrectly bet R	3	2	6	2
	incorrectly bet L	4	9	0	3
White	incorrectly bet R	13	15	29	7
	incorrectly bet L	6	3	0	5
	Total	26	29	35	17

Table 3.14: Probit regressions of quack choices on different measures of test asymmetries. The control variables include the random quack chance for each budget and the interaction point of each pair of budgets.

	Dependent variable: dummy for quack choices					
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	-2.76**	-2.78**	-2.59**	-2.45**	-3.36**	-2.88**
	(0.65)	(0.64)	(0.63)	(0.66)	(0.90)	(0.69)
Quack chance	2.54**	2.52**	2.52**	2.52**	2.77**	2.56**
	(0.26)	(0.26)	(0.26)	(0.26)	(0.33)	(0.26)
Pivot point	1.55	1.49	1.24	1.31	2.66*	1.78*
	(0.84)	(0.83)	(0.85)	(0.86)	(1.30)	(0.88)
$\|p-q\|$	0.18					
	(0.17)					
$\|\|(p,q)-\text{pivot}\|\|$		0.43$\cdot$				
		(0.22)				
$\mathbb{P}$(red)			0.24			
			(0.21)			
q/p				-0.04		
				(0.06)		
$(q-\text{pivot})/(\text{pivot}-p)$					-0.05	
					(0.04)	
$(\mu_l-\mu)/(\mu-\mu_r)$						0.05
						(0.05)
Observations	696	696	696	696	696	696
Note:				$\cdot$p<0.1; *p<0.05; **p<0.01		

Individual comments and decision rules

Table 3.15: Individual quack choice rates, Bayesian updating coefficients, and comments

ID	Quack	$1 - \hat{\alpha}_1$	Comments
4	0.29	0.04	while betting white gives R a small advantage, it also nearly guarantees an L on a red ball, making any choice valid
7	0.21	0.44	I did it where I would have a much better chance at getting a certain number and percentage to chose from
8	0.43	0.11	The colour choices are based on the difference in red and white between L and R, you make the gap as big as possible so its easier to choose L or R from red and white. The bet is then based on the ratio between the gaps.
10	0.29	-0.00	Made the easiest to calculate, for most secure bets.
11	0.29	0.00	If there is only one kind of a color, for ex, all white are L (like in this case), one of the chances is eliminated, as there is no possible bet to do. If the ball is white, it is 100% sure L, so we only bet once in two tries (kind of).
13	0.21	-0.00	My strategy for each round was to try and determine what selection of colors would give the clearest option for the largest number of balls. In the last round, 14, I isolated 70 red balls on the left side and left no red balls on the right side, that way if a ball was red, it was guaranteed to be a left-side ball. This means that I have certainty on a little more than a third of the balls, and then determining if the ball was left or right when the ball was white was closer to a coin flip. I determined my bet choices by simply dividing the number of balls on one side of a color by the total number of balls of that color.
16	0.36	0.00	i just tried to find the optimal split so that the coloured balls would have the most value
18	0.29	0.02	İ tried to make the biggest difference in number of coloured and non-coloured balls
20	0.36	0.00	I was trying to make 100% chance in one of the colours, and count the other probability.
21	0.29	0.24	I tried, if possible, to have a color appear only in one box so as to make sure that it will 100% be in that box. If that wasn't possible, I tried to gather as much of it in one box so as to minimize/maximize the chances.
22	0.50	0.21	I tried to make the ratios of white in box L and R as great as possible

Individual quack choice rates, Bayesian updating coefficients, and comments (cnt)

ID	Quack	$1-\hat{\alpha}_1$	Comments
23	0.29	0.24	Initially, I observe which of the two colours has the safest ratio, in this case I identified RED L as the safest colour to place my bet on. Secondly, I see that the two whites are now slightly higher than a 50/50 chance, so I naturally get on the WHITE R as the odds are ever-so-slightly in my favour, meanwhile I am certain to win the bet on the red colours.
24	0.50	0.004	Tried to do it with an excel spreadsheet but ended up just using basic math knowledge.
26	0.29	-0.00	I choose the colour trying to have the balls in one of the boxes all of the same colour, like all white balls. When this was possible I was able to bet that in that box I had 0% possibilities to finds a red ball. In the other cases I just calculated the percentage mathematically
27	0.29	0.01	I moved the sliders until I thought that the differences in the numbers of white and red balls in box R and L were far enough apart to make a significant difference in statistics (I was just eyeballing it). Once I was satisfied, I confirmed the selection and then worked out the probabilities by taking the number of coloured balls in one box and dividing by the total number of coloured balls to work out the chance of it occurring. E.g. Red L / Red L + Red R = 0.71 so if the ball is red I know there is a 71% chance of it being red. I did this all the time so my bets where always statistically likely to be right (even if just by a little)
28	0.36	0.01	calculated how likely it was to get left or right for both red and white balls and calculated the chance of getting a red or white ball. Did this for both the extremes of colour choices and compared
29	0.43	0.15	i tried making the colours based on which had the furthest overall chances from 50/50 in both categories at the same time, then (mostly) bet on those with higher chances
30	0.21	0.02	tried to make it so only one box had one colour so it would definitely come from that box
35	0.29	0.15	I looked at the ratio between red and white in each box, and make them different
36	0.21	0.04	Rough estimation of possibilities for different outcomes (considering the amount of colored balls) then choosing the one that has the most chance of winning
37	0.29	0.28	Simply Getting as many red balls as possible for L
39	0.29	0.02	I tried to balance the colors in the way that should make it easier for me to guess the chance in the bet choices. When betting, I roughly estimated the percentages looking on the number of balls of given color.

Individual quack choice rates, Bayesian updating coefficients, and comments (cnt)

ID	Quack	$1-\hat{\alpha}_1$	Comments
41	0.14	-0.00	I tried to somewhat increase the difference between the amount of balls between each box
42	0.36	0.69	I tried to pick the colors to make my choice easier. For example if there was no white balls with the label R, then it was easier to make the choice and estimation.
43	0.29	-0.00	Try to favor one colour, increasing the chances for one colour to have a high change to belong to one of the boxes
44	0.36	0.66	My bet obviously hindered on my color choices but when choosing my color choices I just tried to do a variety of outcomes to see how unique I could make each round. I enjoyed testing how moving one slider influences the other so it was a lot of experimenting.
45	0.36	0.05	I just used the highest chance of it being red or white.
51	0.50	-0.00	Went for the safest option, knowing that if it's red I can guarantee that it will be L
56	0.29	0.07	I moved the slider all the way towards whichever side caused the most red balls to be in one of the two boxes and the least red balls to be in hte second box (in example 9, moving the slider all the way to the left caused 86 balls in box L and only 2 balls in box R). I predicted if ball A is red, it will have been drawn from box L from the percentage calculation of 86/88 meaning that 98% of the red balls in box A are from box L. If the ball is white, I added the total amount of white balls (72 in example 9) and predicted that the ball would come from whichever box had more white balls remaining. In example 9, box R had more white balls and the percentage was calculated by doing 38 balls from box R divided as a total percentage which was 38/72 or 53%.
58	0.21	0.05	i was trying to make the probabilities of getting at least one of the if-bets right by eliminating most on my chances to guess wrong. For example in round 3, out of 14 whites there was only one chance of the ball being R.
59	0.21	0.02	I made sure that wherever i could, there was an option that red or white would 100% be label R or L
60	0.71	0.52	Based on the ratio in each box and separately. Both have the same number of balls and same ratio of colours.
61	0.21	-0.09	I like to make the colour that i have 100% chance to win if it will be that colour, i do bet from estimated chances
63	0.21	-0.00	I made color choices to make the odds in my favor for both if bets and the bet choices were based on those odds.
64	0.43	0.28	I tried to make both colours have similar amount of balls but also similar numbers of them balls so there is a chance the likelihood would be correct

3.C.3 Additional analyses on demographics, attitudes, and cognitive abilities

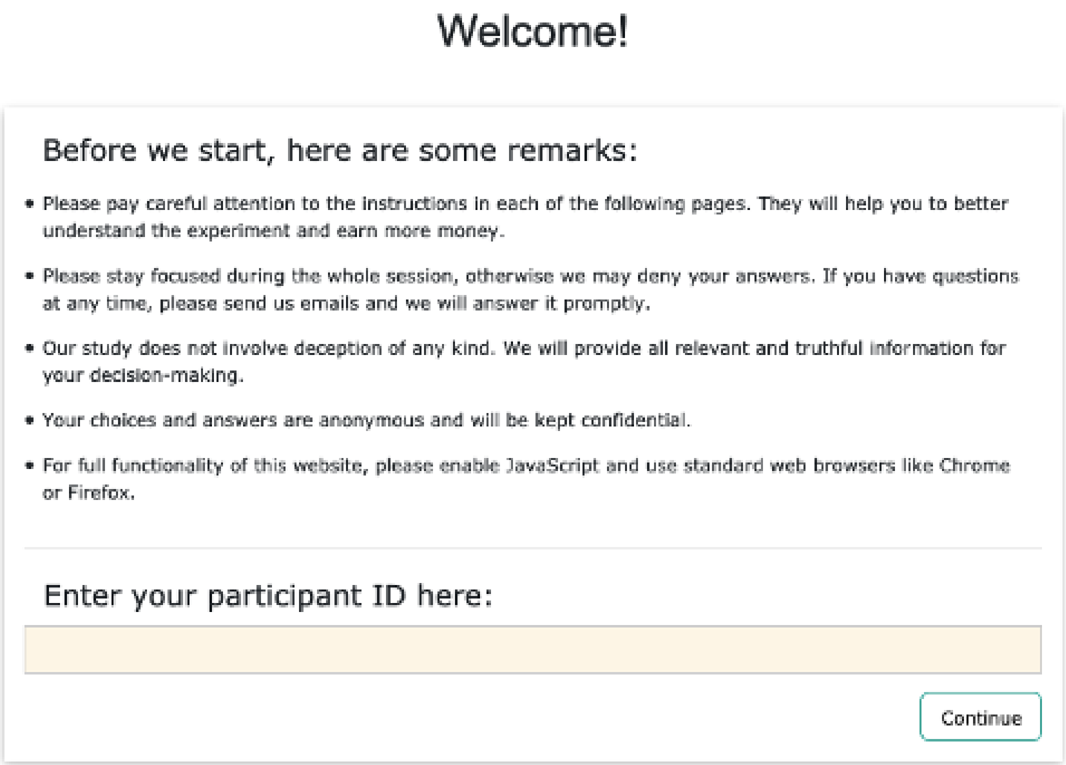
Table 3.16: Regressions of individual quack choice rate on demographic variables, self-reported psychological attitudes, and cognitive reasoning scores. Regressions in columns (4)–(6) are selected based on Akaike (1998)'s information criterion (AIC).

	Dependent variable: individual rate of quacks					
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	0.23	0.20	0.38^{***}	0.14	0.03	0.24
	(0.16)	(0.19)	(0.05)	(0.14)	(0.12)	(0.18)
Age	0.004					
	(0.01)					
Female	0.04					
	(0.04)					
SAT	0.001					
	(0.02)					
STEM	0.02					
	(0.04)					
CRT score			-0.04^{**}	-0.03^{*}	-0.03^{**}	-0.03^{**}
			(0.02)	(0.02)	(0.02)	(0.02)
Wason score			-0.02			
			(0.02)			
Logic score			0.02			
			(0.02)			
Risk		0.04^{**}		0.04^{***}	0.04^{***}	0.04^{**}
		(0.02)		(0.01)	(0.01)	(0.01)
Contingent		-0.04		-0.04		-0.04
		(0.03)		(0.03)		(0.03)
Stubborn		0.02		0.03^{*}	0.03^{*}	0.02
		(0.02)		(0.01)	(0.01)	(0.01)
Information		0.06^{**}		0.05^{**}	0.04^{*}	0.06^{**}
		(0.03)		(0.02)	(0.02)	(0.02)
Perspective		-0.02				-0.02
		(0.03)				(0.03)
Analytical		-0.02				
		(0.02)				
Observations	58	58	58	58	58	58
Adjusted R^2	-0.06	0.15	0.04	0.21	0.19	0.21
Note:				*p<0.1; **p<0.05; ***p<0.01		

Table 3.17: Regressions of individual Bayesian updating bias on demographic variables, self-reported psychological attitudes, and cognitive reasoning scores. Regressions in columns (4)–(6) are selected based on Akaike (1998)'s information criterion (AIC).

	Dependent variable: individual coefficient $1 - \hat{\alpha}_1$					
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	0.03	0.11	0.20^{**}	0.12	0.22	0.07
	(0.25)	(0.38)	(0.09)	(0.20)	(0.19)	(0.21)
Age	-0.004					
	(0.01)					
Female	0.24^{***}			0.21^{***}	0.21^{***}	0.21^{***}
	(0.07)			(0.05)	(0.05)	(0.05)
SAT	0.004					
	(0.03)					
STEM	0.06					
	(0.07)					
CRT score			-0.09^{***}	-0.08^{***}	-0.07^{***}	-0.09^{***}
			(0.03)	(0.03)	(0.03)	(0.03)
Wason score			-0.04			
			(0.03)			
Logic score			0.06			0.04
			(0.04)			(0.03)
Risk aversion		0.03				
		(0.03)				
Contingent		0.01				
		(0.06)				
Stubborn		-0.02				
		(0.03)				
Information		0.03		0.06		0.05
		(0.05)		(0.04)		(0.04)
Perspective		-0.07		-0.10^{**}	-0.08^{**}	-0.09^{**}
		(0.05)		(0.04)	(0.04)	(0.04)
Analytical		0.02		0.06	0.07^{**}	0.06
		(0.05)		(0.04)	(0.04)	(0.04)
Observations	58	58	58	58	58	58
Adjusted R^2	0.17	-0.06	0.13	0.33	0.31	0.33
Note:				$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$		

Appendix 3.D Experimental instructions and screenshots

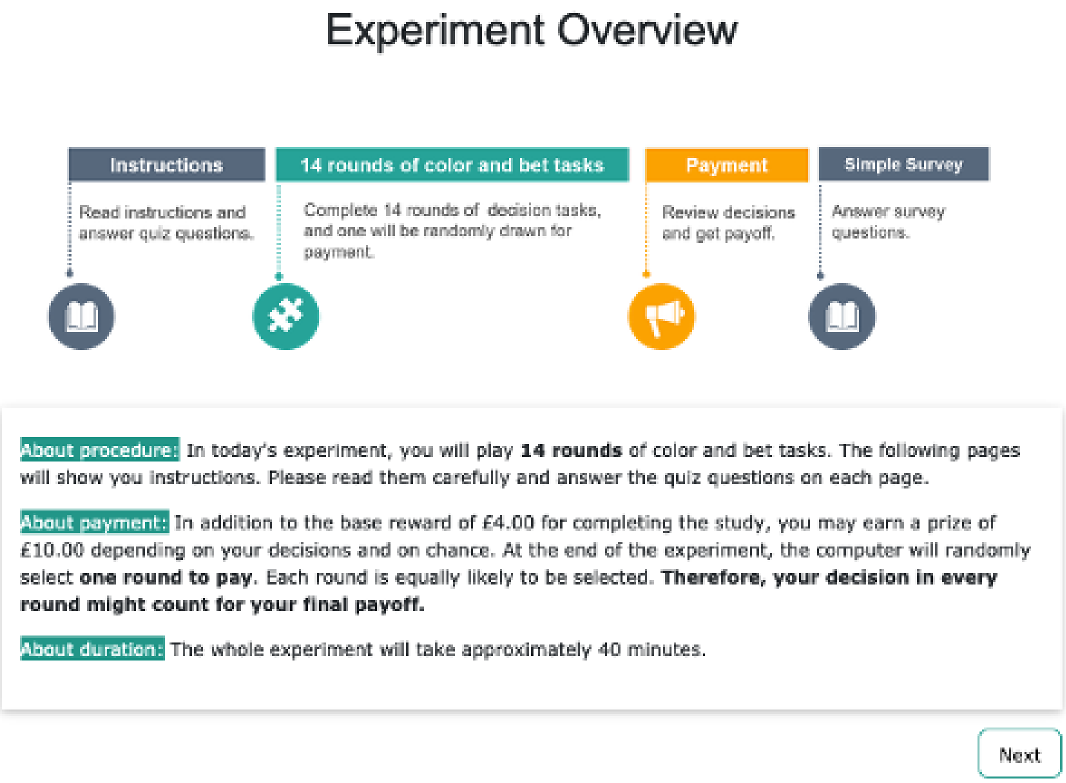


Welcome!

Before we start, here are some remarks:

- Please pay careful attention to the instructions in each of the following pages. They will help you to better understand the experiment and earn more money.
- Please stay focused during the whole session, otherwise we may deny your answers. If you have questions at any time, please send us emails and we will answer it promptly.
- Our study does not involve deception of any kind. We will provide all relevant and truthful information for your decision-making.
- Your choices and answers are anonymous and will be kept confidential.
- For full functionality of this website, please enable JavaScript and use standard web browsers like Chrome or Firefox.

Enter your participant ID here:



Experiment Overview

Instructions

Read instructions and answer quiz questions.

14 rounds of color and bet tasks

Complete 14 rounds of decision tasks, and one will be randomly drawn for payment.

Payment

Review decisions and get payoff.

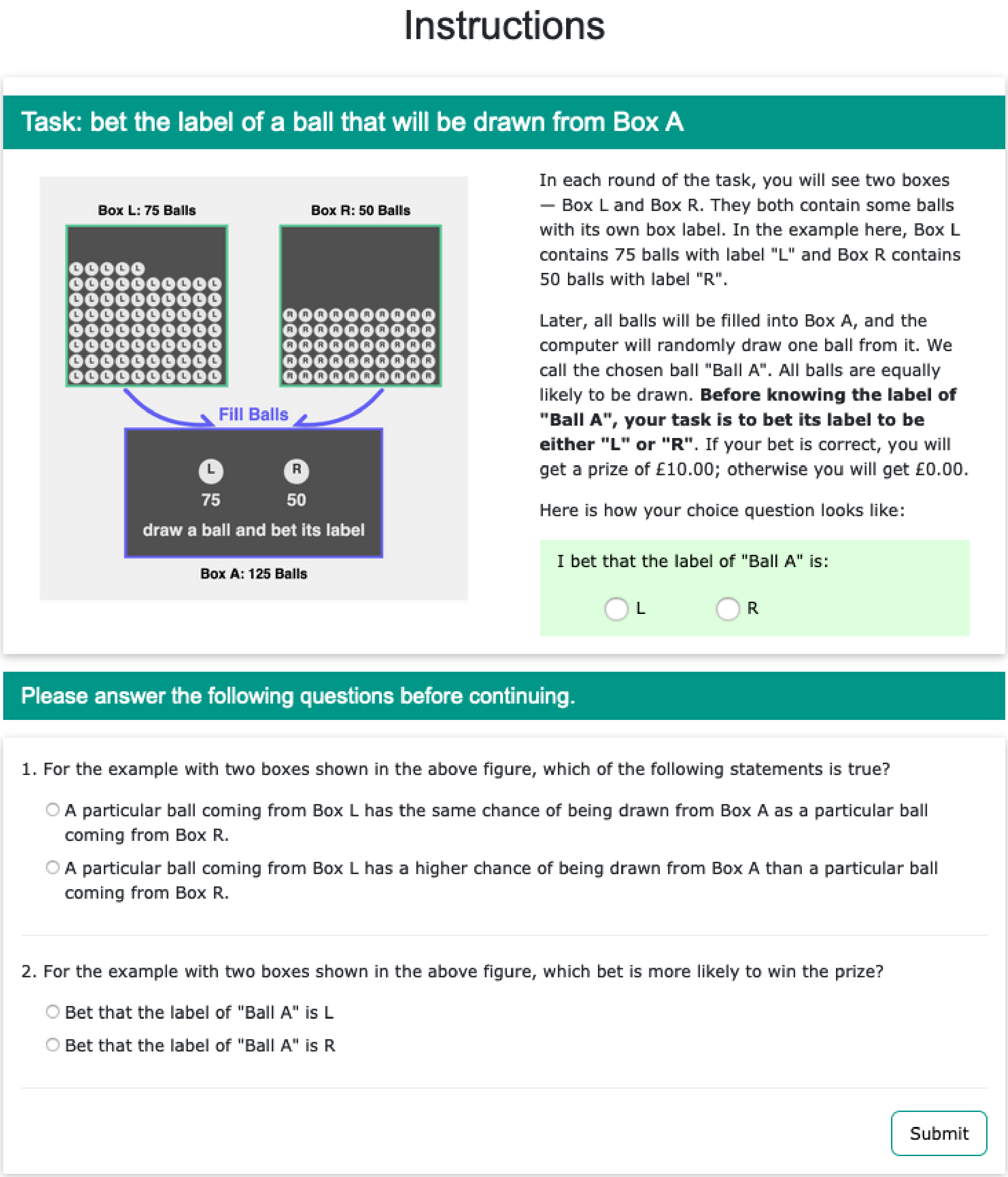
Simple Survey

Answer survey questions.

About procedure: In today's experiment, you will play **14 rounds** of color and bet tasks. The following pages will show you instructions. Please read them carefully and answer the quiz questions on each page.

About payment: In addition to the base reward of £4.00 for completing the study, you may earn a prize of £10.00 depending on your decisions and on chance. At the end of the experiment, the computer will randomly select **one round to pay**. Each round is equally likely to be selected. **Therefore, your decision in every round might count for your final payoff.**

About duration: The whole experiment will take approximately 40 minutes.



Instructions

Task: bet the label of a ball that will be drawn from Box A

In each round of the task, you will see two boxes — Box L and Box R. They both contain some balls with its own box label. In the example here, Box L contains 75 balls with label "L" and Box R contains 50 balls with label "R".

Later, all balls will be filled into Box A, and the computer will randomly draw one ball from it. We call the chosen ball "Ball A". All balls are equally likely to be drawn. **Before knowing the label of "Ball A", your task is to bet its label to be either "L" or "R"**. If your bet is correct, you will get a prize of £10.00; otherwise you will get £0.00.

Here is how your choice question looks like:

I bet that the label of "Ball A" is:

○ L ○ R

Please answer the following questions before continuing.

1. For the example with two boxes shown in the above figure, which of the following statements is true?

○ A particular ball coming from Box L has the same chance of being drawn from Box A as a particular ball coming from Box R.

○ A particular ball coming from Box L has a higher chance of being drawn from Box A than a particular ball coming from Box R.

2. For the example with two boxes shown in the above figure, which bet is more likely to win the prize?

○ Bet that the label of "Ball A" is L

○ Bet that the label of "Ball A" is R

Instructions

Task 1: Choose color compositions:

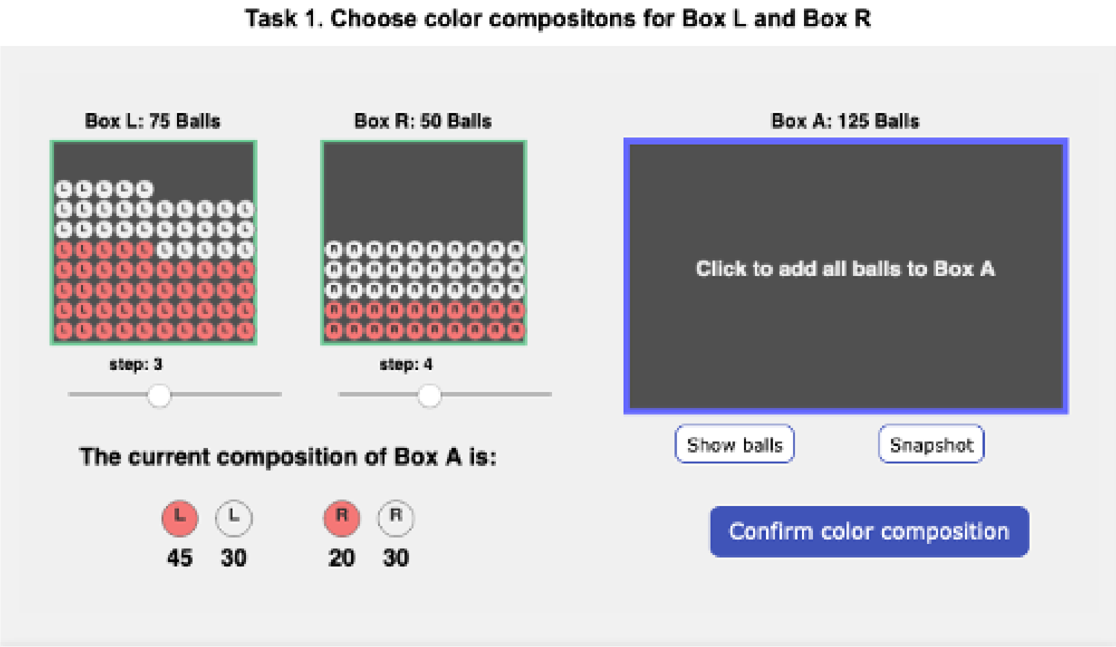
Before putting all balls into Box A, you can choose to color some balls in Box L and Box R to be red. **The coloring process will not affect the label of each ball.**

After choosing the color, all balls in Box L and Box R will be put into Box A, and **you can still bet the label of "Ball A" to win a prize of £10.00.** However, you can make bets separately as if knowing the color of "Ball A".

The coloring task might help you guess the label of "Ball A" better. Why? First, notice that the color composition you choose will determine the likelihood of "Ball A" being red or white. Second, different color compositions for Box L and Box R will help you to infer the label of "Ball A".

How to color balls in Box L and Box R? Below is an example of the choice interface. Play with it to find out!

An example of the choice interface



Task 1. Choose color compositons for Box L and Box R

The current composition of Box A is:

L (red)	L (white)	R (red)	R (white)
45	30	20	30

How to choose color compositions?

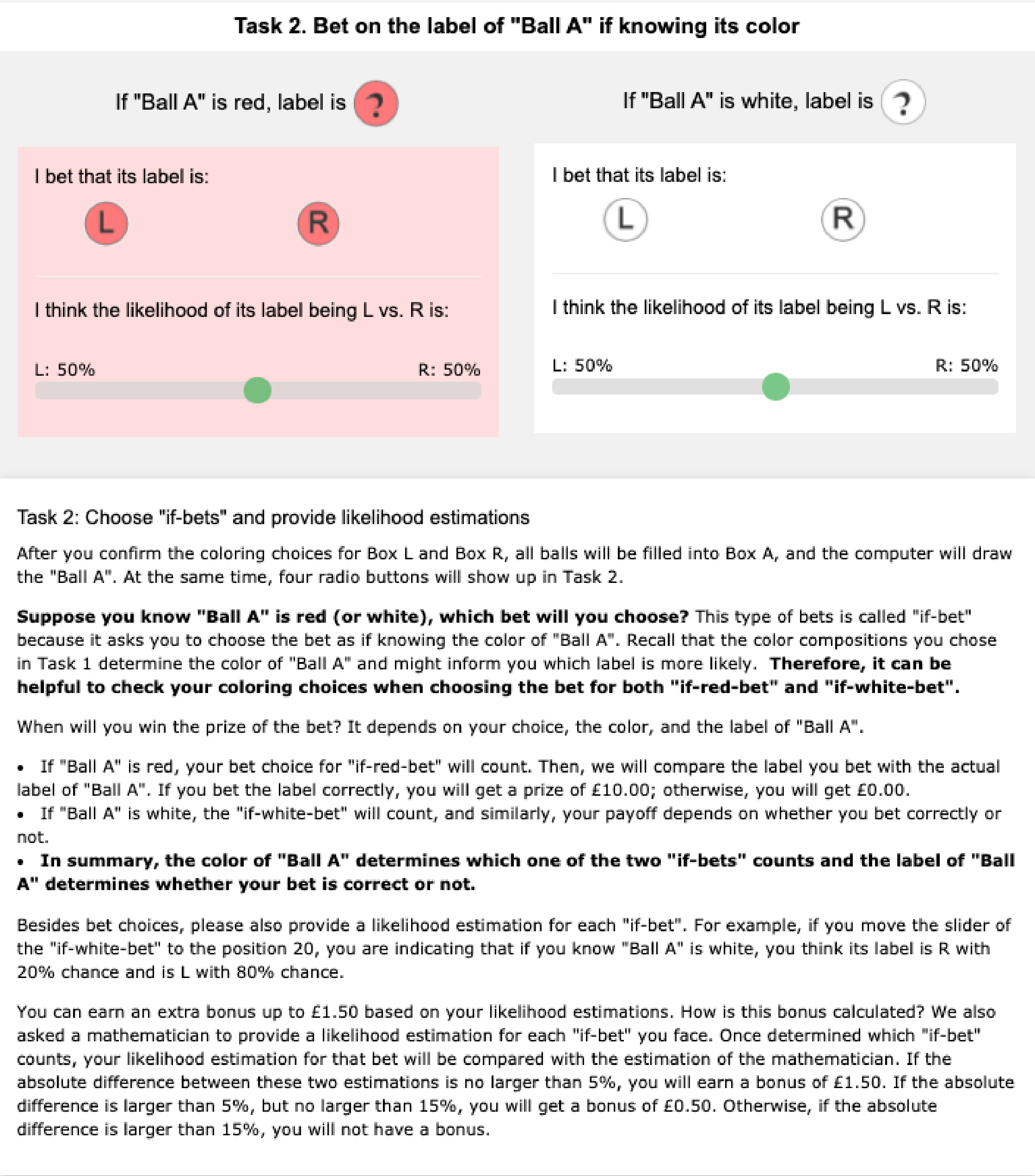
There are two sliders below Box L and Box R. You can directly drag them or use arrow keys on your keyboard to change the number of red balls and white balls in each box.

Moving the slider to the right (left) will increase the number of red (white) balls.

Notice that **the increments for red (white) balls in Box L and Box R are linked.** In the example here, if you add three red (white) balls for Box L, four red (white) balls will be automatically added to Box R. The increment steps for each box is shown above the slider.

You can also check how the colored and labeled balls look like in Box A by clicking two buttons below Box A.

Once you are satisfied with your coloring choice, please confirm it by clicking the button "Confirm color composition". After that, you cannot change the color of the balls anymore. In actual tasks, two sliders will disappear.



Task 2. Bet on the label of "Ball A" if knowing its color

If "Ball A" is red, label is ?

I bet that its label is:

○ L ○ R

I think the likelihood of its label being L vs. R is:

L: 50% R: 50%

If "Ball A" is white, label is ?

I bet that its label is:

○ L ○ R

I think the likelihood of its label being L vs. R is:

L: 50% R: 50%

Task 2: Choose "if-bets" and provide likelihood estimations

After you confirm the coloring choices for Box L and Box R, all balls will be filled into Box A, and the computer will draw the "Ball A". At the same time, four radio buttons will show up in Task 2.

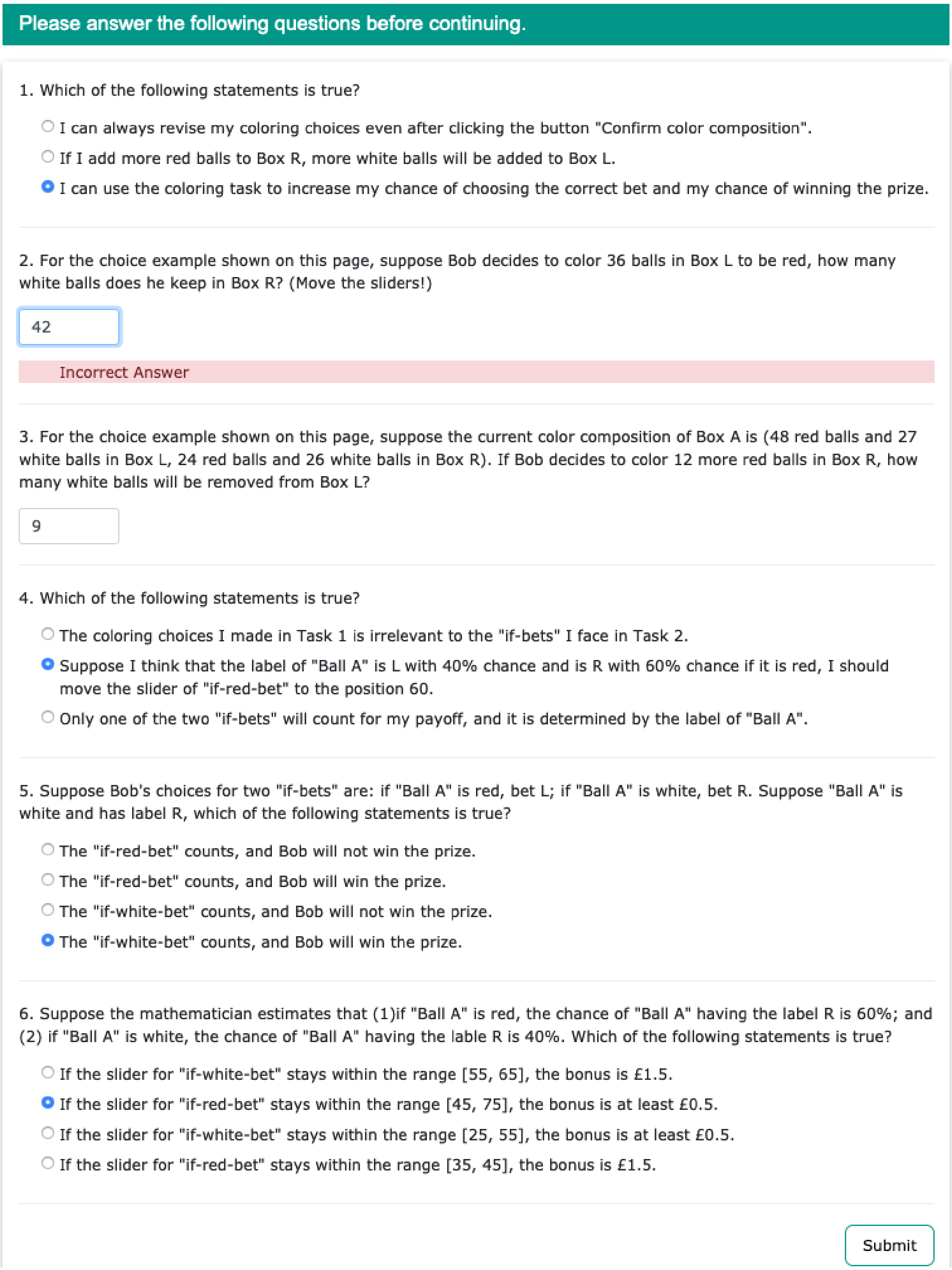
Suppose you know "Ball A" is red (or white), which bet will you choose? This type of bets is called "if-bet" because it asks you to choose the bet as if knowing the color of "Ball A". Recall that the color compositions you chose in Task 1 determine the color of "Ball A" and might inform you which label is more likely. **Therefore, it can be helpful to check your coloring choices when choosing the bet for both "if-red-bet" and "if-white-bet".**

When will you win the prize of the bet? It depends on your choice, the color, and the label of "Ball A".

- If "Ball A" is red, your bet choice for "if-red-bet" will count. Then, we will compare the label you bet with the actual label of "Ball A". If you bet the label correctly, you will get a prize of £10.00; otherwise, you will get £0.00.
- If "Ball A" is white, the "if-white-bet" will count, and similarly, your payoff depends on whether you bet correctly or not.
- **In summary, the color of "Ball A" determines which one of the two "if-bets" counts and the label of "Ball A" determines whether your bet is correct or not.**

Besides bet choices, please also provide a likelihood estimation for each "if-bet". For example, if you move the slider of the "if-white-bet" to the position 20, you are indicating that if you know "Ball A" is white, you think its label is R with 20% chance and is L with 80% chance.

You can earn an extra bonus up to £1.50 based on your likelihood estimations. How is this bonus calculated? We also asked a mathematician to provide a likelihood estimation for each "if-bet" you face. Once determined which "if-bet" counts, your likelihood estimation for that bet will be compared with the estimation of the mathematician. If the absolute difference between these two estimations is no larger than 5%, you will earn a bonus of £1.50. If the absolute difference is larger than 5%, but no larger than 15%, you will get a bonus of £0.50. Otherwise, if the absolute difference is larger than 15%, you will not have a bonus.



Please answer the following questions before continuing.

1. Which of the following statements is true?

○ I can always revise my coloring choices even after clicking the button "Confirm color composition".

○ If I add more red balls to Box R, more white balls will be added to Box L.

● I can use the coloring task to increase my chance of choosing the correct bet and my chance of winning the prize.

2. For the choice example shown on this page, suppose Bob decides to color 36 balls in Box L to be red, how many white balls does he keep in Box R? (Move the sliders!)

42

Incorrect Answer

3. For the choice example shown on this page, suppose the current color composition of Box A is (48 red balls and 27 white balls in Box L, 24 red balls and 26 white balls in Box R). If Bob decides to color 12 more red balls in Box R, how many white balls will be removed from Box L?

9

4. Which of the following statements is true?

○ The coloring choices I made in Task 1 is irrelevant to the "if-bets" I face in Task 2.

● Suppose I think that the label of "Ball A" is L with 40% chance and is R with 60% chance if it is red, I should move the slider of "if-red-bet" to the position 60.

○ Only one of the two "if-bets" will count for my payoff, and it is determined by the label of "Ball A".

5. Suppose Bob's choices for two "if-bets" are: if "Ball A" is red, bet L; if "Ball A" is white, bet R. Suppose "Ball A" is white and has label R, which of the following statements is true?

○ The "if-red-bet" counts, and Bob will not win the prize.

○ The "if-red-bet" counts, and Bob will win the prize.

○ The "if-white-bet" counts, and Bob will not win the prize.

● The "if-white-bet" counts, and Bob will win the prize.

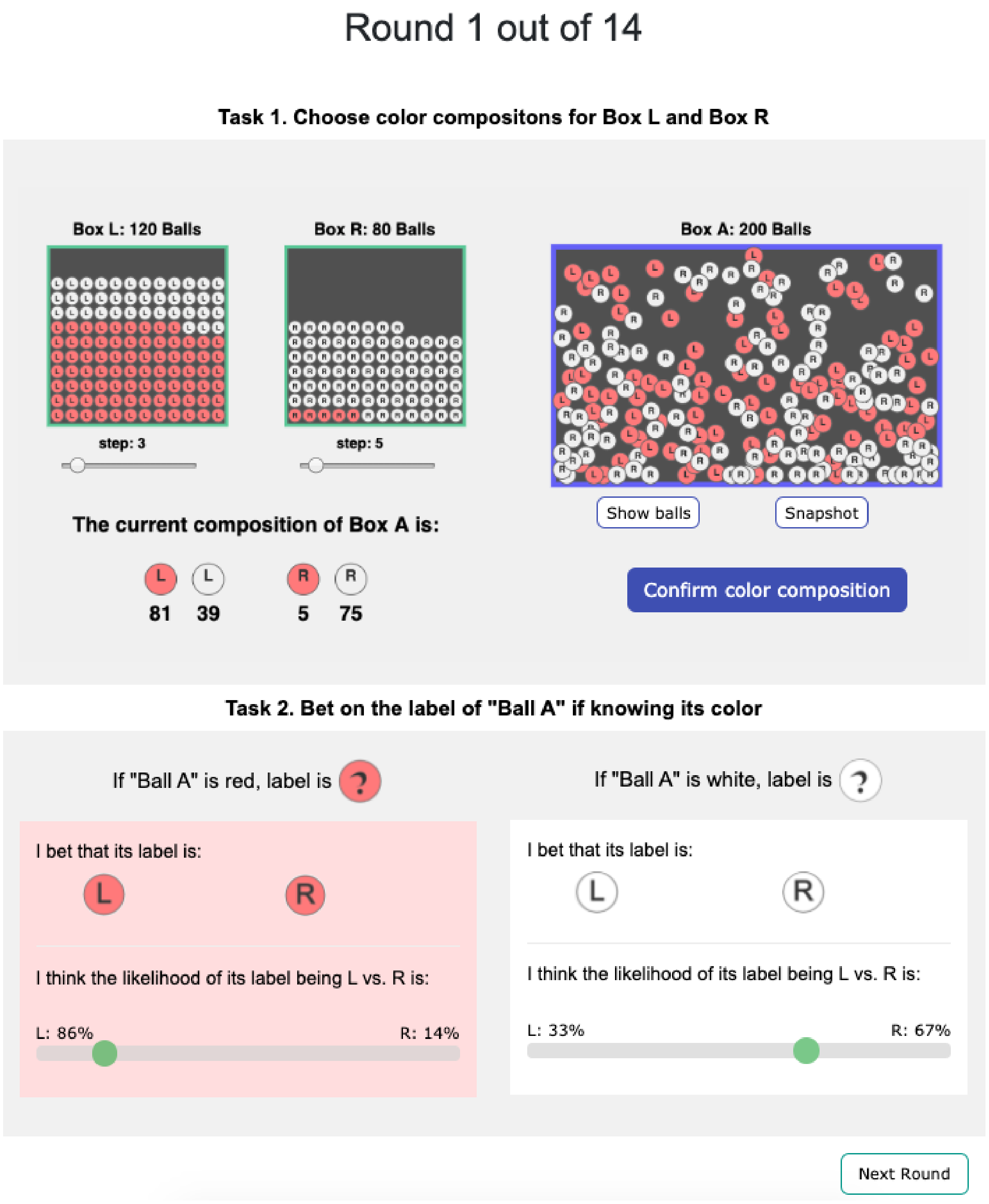
6. Suppose the mathematician estimates that (1)if "Ball A" is red, the chance of "Ball A" having the label R is 60%; and (2) if "Ball A" is white, the chance of "Ball A" having the lable R is 40%. Which of the following statements is true?

○ If the slider for "if-white-bet" stays within the range [55, 65], the bonus is £1.5.

● If the slider for "if-red-bet" stays within the range [45, 75], the bonus is at least £0.5.

○ If the slider for "if-white-bet" stays within the range [25, 55], the bonus is at least £0.5.

○ If the slider for "if-red-bet" stays within the range [35, 45], the bonus is £1.5.



Round 1 out of 14

Task 1. Choose color compositons for Box L and Box R

Box L: 120 Balls

Box R: 80 Balls

Box A: 200 Balls

step: 3

step: 5

The current composition of Box A is:

L (red)	L (white)	R (red)	R (white)
81	39	5	75

Task 2. Bet on the label of "Ball A" if knowing its color

If "Ball A" is red, label is ?

I bet that its label is:

I think the likelihood of its label being L vs. R is:

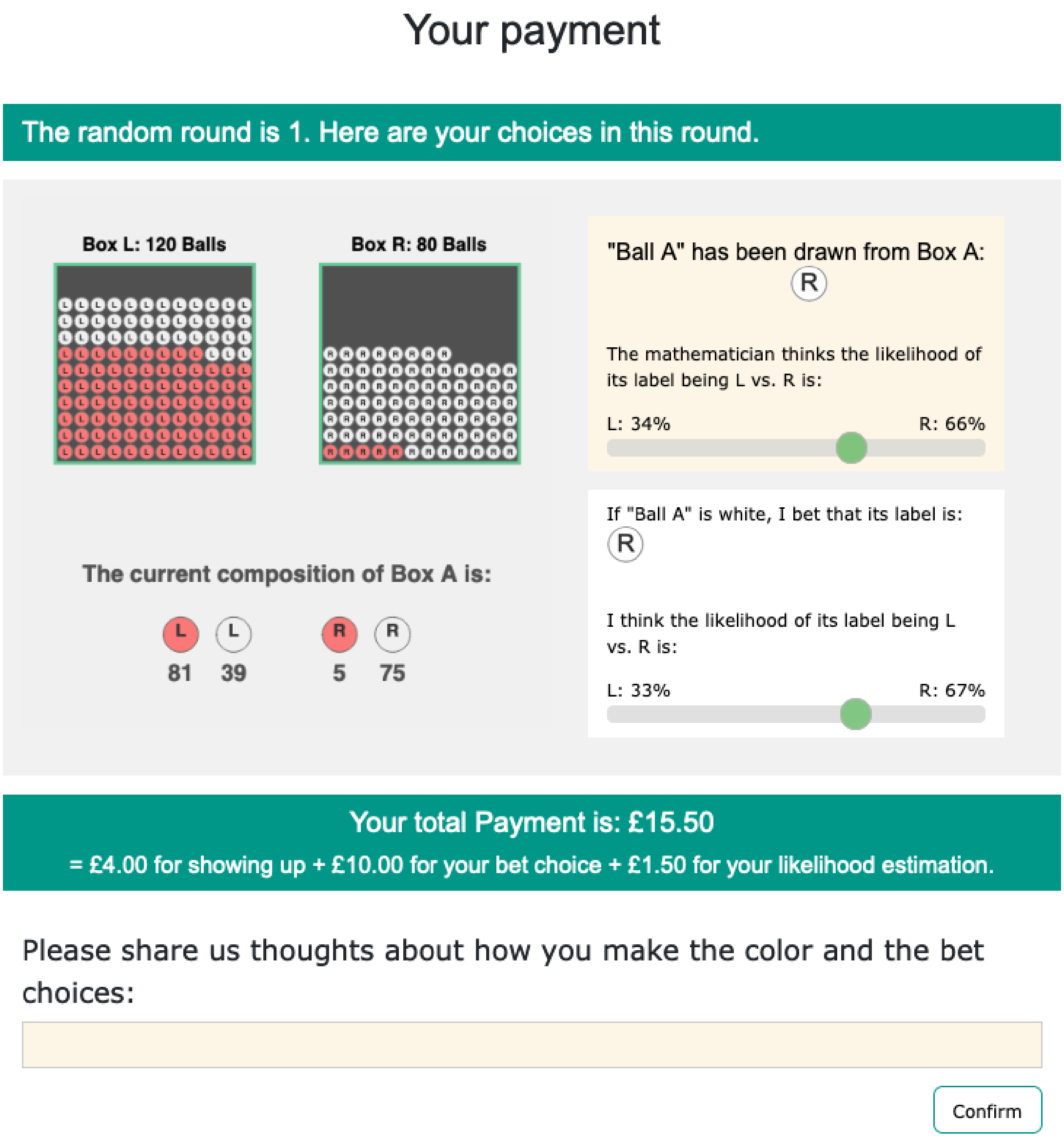
L: 86% R: 14%

If "Ball A" is white, label is ?

I bet that its label is:

I think the likelihood of its label being L vs. R is:

L: 33% R: 67%



Your payment

The random round is 1. Here are your choices in this round.

The current composition of Box A is:

L	L	R	R
81	39	5	75

"Ball A" has been drawn from Box A: R

The mathematician thinks the likelihood of its label being L vs. R is:

L: 34% R: 66%

If "Ball A" is white, I bet that its label is: R

I think the likelihood of its label being L vs. R is:

L: 33% R: 67%

Your total Payment is: £15.50

= £4.00 for showing up + £10.00 for your bet choice + £1.50 for your likelihood estimation.

Please share us thoughts about how you make the color and the bet choices:

Chapter 4

Will Bayesian markets induce truth-telling? —An experimental study

Abstract: The Bayesian market (Baillon (2017)) is a new mechanism that incentivizes individuals to report their private signals truthfully. This chapter tests the performance of Bayesian markets and studies how they are influenced by the equilibrium requirement of truth-telling. I construct laboratory Bayesian markets with three different degrees of manipulation in participants' beliefs over others' truthfulness. I find that Bayesian markets effectively induce truthful revelations when participants believe that others are truthful. However, when there are noises in agents' beliefs, Bayesian markets become less effective. The existence of speculative buyers in the market exacerbates participants' under-inference bias in processing private information. In the market with the most significant disturbances, individuals ignore their private signals. As a result, they expect the value of the asset is higher than its fundamental value and thus are more likely to buy an asset. The over-buying inclination raises the ex-post realization of asset values in Bayesian markets, leading to market bubbles and under-performance of the mechanism.

4.1 Introduction

Information elicited from dispersed individuals is increasingly crucial to many knowledge-gathering and decision-making tasks. Researchers are conducting extensive social-economic surveys to collect knowledge about human perceptions. Crowdsourcing platforms are keen to solicit informative answers from online communities. Customers are steering their purchases to products with reliable customer reviews. In most cases, information providers are compensated for their time and efforts, if at all, not for the quality of their answers; consequently, they may provide uninformative or untruthful answers. A worker who labels images on MTurk[1], for instance, may type random tags given the wage is fixed. A survey-taker may lie about sensitive questions involving drug abuse or impaired driving. When the answers from the crowd are unreliable, the algorithms and decisions based on them will be biased. How to encourage respondents to report their private information truthfully is the biggest challenge in information elicitation.

When a verifiable truth is present to condition payment, many mechanisms create incentives for truth-telling. Consider proper scoring rules (Winkler (1969)).[2] A respondent submits a probabilistic forecast of an event, and then the report is graded against the objective truth – the realization of the event or its frequency – by a score. Typically, the closer a report is to the underlying truth, the higher the score it will attain, and correspondingly the higher the reward will be (see Gneiting and Raftery (2007) and Gneiting and Katzfuss (2014)). The prediction market is another popular elicitation mechanism.[3] A participant buys or sells a contract paying one dollar if an event occurs. Wolfers and Zitzewitz (2006) showed that the asset price on a prediction market is close to the mean belief of all participants. Hence, an agent whose prediction probability of the event is higher than the average belief will buy a contract. An agent whose prediction is lower than the average belief will sell a contract.

Observing the objective truth is crucial for scoring rules and prediction markets in incentivizing truth-telling. In the example of scoring rules, the event outcome or its actual frequency serves as an evaluation gauge for all possible forecasts from individuals. In the presidential prediction market, the election outcome verifies profits for all contracts. Both mechanisms condition individuals' payoffs with the observable outcome of the underlying event. Therefore, truthful reports of their opinions and beliefs of the event yield the highest expected payoff. Principally, the validity of an elicitation mechanism hinges upon how well the monetary incentives are aligned to informants' truthful revelations.

[1] MTurk (`https://www.mturk.com`) is a crowd-sourcing platform enabling individuals to perform Human Intelligence Tasks to earn money.

[2] Early work also include Brier (1950), Good (1950), Winkler and Murphy (1968), and Savage (1971). Schlag, Tremewan and Van der Weele (2015) compared various scoring rules. Offerman et al. (2009) generalized them for non-expected utilities. Schotter and Trevino (2014) reviewed their empirical implementations.

[3] Early inspirations of prediction market are from Hayek (1945) and Fama (1970)'s discussions on the efficient information aggregation in financial markets. Related reviews include Manski (2006), Tziralis and Tatsiopoulos (2007), Berg et al. (2008), and Berg, Nelson and Rietz (2008). Recent applications can be found in Arrow et al. (2008), Dreber et al. (2015), Camerer et al. (2016), and Camerer et al. (2018). In practice, the Iowa Electronic Markets (`https://iemweb.biz.uiowa.edu/`) are running examples of collecting individual opinions on political elections through market transactions.

When the underlying truth is subjective or costly to verify, it is less intuitive to apply the incentive alignment principle. For questions like “Have you ever engaged in questionable research practice?”, “Do you believe computers will outsmart humans?” or “Are you happy?”, there is no natural benchmark against which respondents' answers are evaluated, and truthful ones are rewarded. Nevertheless, subjective truths like feelings, judgments, and emotions are prominent in modern life.

Miller, Resnick and Zeckhauser (2005)'s peer prediction method and Prelec (2004)'s Bayesian truth serum (BTS) are two classes of mechanisms eliciting the unverifiable truth.[4] They exploit the relationship of private information among the population and construct benchmarks for respondents' answers based on peers' answers. In the example regarding the survey question, “Are you happy?”, the implementer of the peer prediction mechanism (called a center) is assumed to know the prior distribution of private information. She transforms each respondent's answer into a belief about a reference's answer through Bayesian updating and further evaluates the belief through a proper scoring rule. BTS relaxes the assumption of the center knowing the prior distribution. Each participant answers“yes” or “no” to the question and also predicts the proportion of participants in a large population answering “yes.” The center assigns each participant a prediction score and an information score. The prediction score induces truthful prediction of the distribution of signal reports. The information score exploits the implied Bayesian reasoning about population frequencies and rewards truthful signal reports that are more common than collectively predicted ones. Since peers' answers are verifiable, truth-telling proves to be a Bayesian Nash Equilibrium for both mechanisms.

The practical validity of peer predictions and Bayesian truth serums remains an open question. Gao et al. (2014) experimentally tested the peer prediction method on MTurk. They found players were more likely to coordinate at uninformative equilibria, suggesting a failure of peer prediction in inducing the truth-telling equilibrium. John, Loewenstein and Prelec (2012) employed BTS incentive schemes and surveyed 2000 psychologists on their involvement in questionable research practices. They found BTS induced a higher self-admission rate. Weaver and Prelec (2013) tested BTS in a recognition questionnaire containing foil brand names or scientific terms. They showed that participants claimed to recognize fewer foils in BTS groups than in control groups, further supporting BTS's capability in inducing truth-telling. In general, it is very difficult to explain to subjects how these score-based mechanisms incentivize truthful

[4] Jurca and Faltings (2006, 2009) extended the peer prediction method in avoiding uninformative equilibria. Parkes and Witkowski (2012) proposed Robust BTS for small population. Radanovic and Faltings (2013, 2014) generalized RBTS to non-binary and continuous signals.

answers. Indeed, subjects in these experiments could not link their actions to payoffs directly, and they were suggested to believe that truth-telling was in their best interests. Suspicions and demand effects may arise. Shaw, Horton and Chen (2011) employed BTS as a contextual manipulation on MTurk and found that workers performed significantly better, even though they were not financially rewarded by BTS. They argued that the out-performance of BTS might be attributed to confusion and cognitive demand.

Baillon (2017) proposed a new institution, called Bayesian market, to simplify the practical implementation of BTS. Through a market where private information is linked to asset transactions, individuals are rewarded for the truthful revelation of their subjective truths. A Bayesian market associated with the question "Are you happy?" works as follows: First, each agent has an opportunity to participate in the market; he can buy (sell) at most one asset by submitting a "yes" ("no") report regarding the question. Then, the market price of the asset is randomly drawn, and agents are asked whether they would like to trade at the price. The asset value is the realized proportion of people answering "yes" to this question. Similar to the Bayesian reasoning in peer prediction and BTS mechanisms, agents who are truly happy expect a higher proportion of "yes" report and thus have a higher valuation of the asset. Hence, they are more likely to buy an asset than those who feel unhappy. Under mild assumptions, Bayesian markets predict a truth-telling BNE.

This paper tests the validity of Bayesian markets in inducing truth-telling. Compare to score-based mechanisms, agents are more familiar with asset transactions and thus may be more engaged in Bayesian markets. One obstacle for the empirical performance of Bayesian markets is the possible confusion and cognitive demands. Another one inherits in the definition of truth-telling BNE: the truthful report is optimal for an agent if he believes all other agents are truthful. Moreover, the belief disturbances in others' truthfulness might be the main source of cognitive demands. To understand how belief disturbances influence the effectiveness of Bayesian markets, I manipulate individual beliefs over others' truthfulness with varying extents. In particular, I study the following three questions: (1) Will Bayesian markets induce the best response of truth-telling from individuals when they believe all others are truthful? (2) Will Bayesian markets induce the best response of truth-telling when participants expect that some agents may lie? (3) Will Bayesian markets induce the truth-telling BNE?

I answer these questions by constructing three types of laboratory Bayesian markets, each featuring a setting where participants' beliefs over others' truthfulness are distorted to a certain degree. In the experiment, agents' signals are linked through two types of bingo cages. The number of each type implies the common prior, and each one of the cages may be the source to generate private signals for all agents. After drawing a signal, each agent updates beliefs over the asset value, decides his trading position in the market, and submits a bid or ask price. The distortion of beliefs is achieved by Algorithm Agents (AAs), who always report their private signals and correct bids/asks under Bayesian posteriors. By varying proportions of AAs among the total eight agents in the market, I expose human agents (HAs) to different degrees of beliefs over others' truthfulness. I consider three treatments: 1HA, 3HA, and 8HA treatment. Each one aims to address one research question proposed before.

I find Bayesian markets effectively induce truthful reports of private signals and posterior expectations of the asset value when the belief system is perfect. When there are disturbances of agents' beliefs over others' truthfulness, Bayesian markets are less effective. The treatment differences are explained by the rise of bubbles in the market. First, speculative buyers are more active in the market with more noises in participants' belief systems. Their existence stimulates agents' under-inference biases in updating private signals. Consequently, agents ignore private signals and chase the trend in the market. They over-estimate the value of the assets and are more likely to buy than short sell assets. Furthermore, the ex-ante expectation of the asset value is self-confirmed by its ex-post realization, and thus bubbles arise.

The rest of the chapter is structured as follows. Section 4.2 describes how a Bayesian market works and its theoretical predictions. Section 4.3 introduces the experiment. In section 4.4 and 4.5, I analyze the performances of Bayesian markets and subjects' transaction decisions. Section 4.6 concludes and discusses potential future works.

4.2 Bayesian market mechanism

4.2.1 Model setup

Consider the simplest case that n homogeneous risk-neutral agents are surveyed with the same binary question — "Are you happy?". The population size n is assumed to be infinite for theoretical simplicity in this section but will be relaxed in the experiment. Agent i's private signal is assumed to be a random variable $T_i \in \{Y, N\}$. Before receiving a realization of the private signal, agent i believes that signals are drawn from a joint distribution $f(T_1, T_2, ..., T_n)$. This prior distribution is assumed to be common knowledge among all agents. Under some regular assumptions, the common prior over signals implies a common prior $f(\omega)$, where ω is the proportion of agents whose private signal is Y.[5] Given a signal realization denoted as t_i, agent i updates his belief of ω through $f(\omega \mid T_i = t_i)$.

A Bayesian market is set up to elicit individuals' private signals. Each participant can trade at most one asset by submitting a yes/no report to the question of interest. The value of the asset v is the proportion of agents reporting yes. The asset's market price, denoted as p, is randomly drawn from a commonly known uniform interval $[0,1]$.[6] Individuals' reports determine their trading positions on the market. After knowing the price, each agent answers whether he would like to buy an asset at p if he reported "yes", and whether he would like to sell an asset at p if he reported "no". There is a market maker ("she") between buyers and sellers. Agents do not trade with each other but with the market maker. This procedure is to prevent agents from learning private information from each other. She executes a transaction for a buyer (seller) if a majority of the agents reporting "yes" ("no") is willing to buy (sell).

Agents' trading decisions rely on their private signals and how they evaluate the asset given different signals. To better understand how disturbances in belief systems influence people's posterior valuations of the asset, I deviate from Baillon (2017) and implement the BDM method (Becker, DeGroot and Marschak (1964)) to elicit people's posterior willingness to pay/accept. Before knowing the market price, each agent submits the highest price he would like to buy an asset (a bid $b_i \in [0,1]$) if he reported "yes", and the lowest price he would like to sell an asset (an ask price $a_i \in [0,1]$) if he reported "no". The market maker collects all bids and asks from participants and uses the average bid and ask price as her buying ($\bar{a}$) and selling ($\bar{b}$) price. The trading rule resembles the majority rule. A trade occurs for buyer i if $b_i \geq p \geq \bar{a}$, under which both the buyer and the market maker are willing to trade at the random price p. Similarly, a seller will successfully short sell an asset to the market maker if $a_i \leq p \leq \bar{b}$.[7]

After a market closes, the market maker will liquidate each asset at its settlement value v. Trading buyers will receive an amount of money equaling v, and trading sellers need to pay back an asset at the cost of v. Hence, the profit is $v - p$ for a trading buyer and $p - v$ for a trading seller. Those who fail to trade receive zero.

[5] See footnote 19 of Prelec (2004).

[6] Other distributions are also valid. In the experiment, I use the truncated normal distribution to increase the chance of transactions.

[7] This procedure was based on an earlier draft of Baillon (2017).

4.2.2 Theoretical predictions of Bayesian markets

I summarize the theoretical predictions on how agents trade in Bayesian markets and how the market transactions result in truth-telling equilibrium in this subsection. Formal proofs can be found in Baillon (2017) and Baillon (2016).

A Bayesian market's validity in truth-inducing relies on the link between an agent's private signal and his expectation of others' signals. Before receiving private signals, an agent i forms an expectation of Y-type in population according to the common prior $f(\omega)$. After receiving a private signal t_i, agent i updates the expectation based on $f(\omega \mid t_i)$. Signals are assumed to be "impersonally informative". By "informative", the signals provide information about population frequency ω; by "impersonal", agents who receive the same signals will learn in the same way of ω. Mathematically speaking, $f(\omega \mid t_i) = f(\omega \mid t_j) \Leftrightarrow t_i = t_j$ implies both "informative" and "impersonal" property.

The signal structure and the implied Bayesian reasoning form the basis for incentive alignment in Bayesian markets. When receiving different private signals, agents formulate different expectations of others' private information. In particular, agents who are truly happy (Y-type) will expect more happy people in the population than those who are not (N-type). Namely, $E(\omega \mid t_Y) > E(\omega \mid t_N)$. Notice that it is a relative comparison. Both types of agents may expect a minority of N-type in population, however, by exploiting the information in their true answers, happy agents still expect that Y-type signals are more common in the population than what unhappy agents expect.

Prediction 4.1. *The agent who participates in a Bayesian market will submit his posterior expectation of asset value $E(v \mid t_i)$ as his bid or ask price.*

Buyers and sellers in the market submit bids and asks that maximize their posterior expected payoffs. The optimal bids and asks will be $E(v \mid t_i)$. In particular, contingent on agent i's buy/sell decision, his willingness to pay for an asset is $b_i = E(v \mid t_i)$ and his willingness to accept an asset is $a_i = E(v \mid t_i)$. This prediction is not surprising due to the implementation of the BDM method. When market prices are randomly determined, it is incentive compatible for an agent to report his underlying willingness to pay/accept, which equals to the posterior expectation of the asset value. For notation simplicity, I denote $E(v \mid t_Y) \equiv \omega_Y$ and $E(v \mid t_N) \equiv \omega_N$ as the posterior expectation for a Y-type and a N-type agent, respectively. If all agents answer the optimal bids and asks, the market maker's buying price will be $\bar{b} = \omega_Y$, and her selling price will be $\bar{a} = \omega_N$.

Prediction 4.2. *Truth-telling is a BNE in Bayesian markets.*

According to the definition of BNE, when all other agents truthfully report their private signals, the asset value is the same as the frequency of the Y-type signal (ω). By

Prediction 4.1, active agents in the market will report their posterior expectations of asset value, producing ω_Y and ω_N as a market maker's buying price and selling price, respectively. A Y-type agent expects a higher asset value than an N-type agent and thus is willing to buy an asset. Similarly, an N-type agent is willing to sell an asset. In equilibrium, both types of agents will participate in markets and truthfully reveal their private signals through their buy/sell decisions.

4.3 Experimental design and procedures

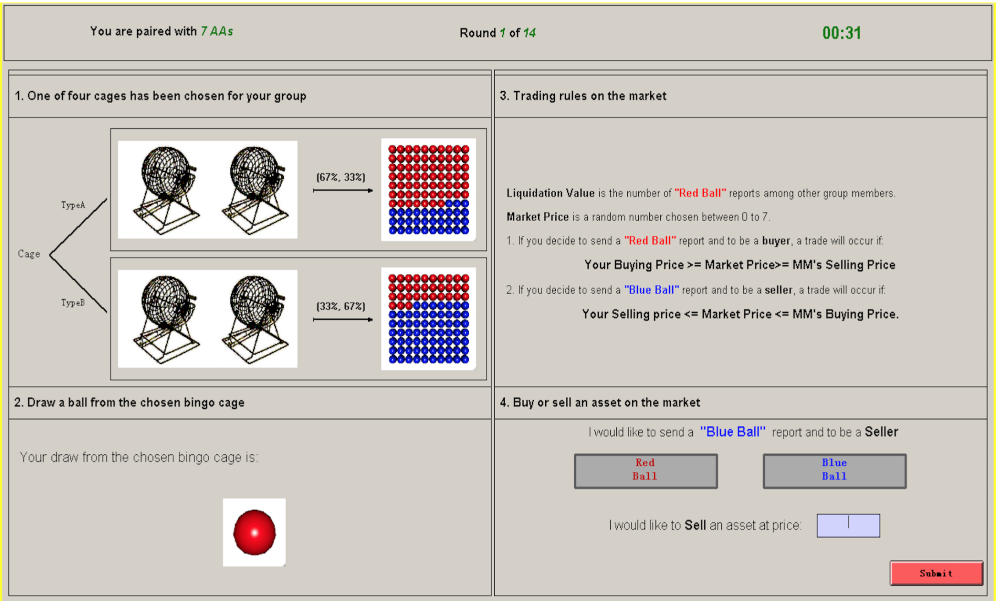
4.3.1 Market structure

The validity of Bayesian markets in inducing truth-telling relies on two critical assumptions — common prior and "impersonally informative" signals. Both are directly defined in the signal space. Unlike mechanisms like peer predictions, Bayesian markets do not require to know the formulation structure of private signals. In practice, it yields convenience. For example, agents may not know the states of the world and how different states generate the signal of being happy or not. However, in some scenarios, such as writing customer reviews, agents form opinions after a realization of the state. Since this paper focuses on whether and how Bayesian markets induce truth-telling, I design the experiment in accordance with the case where both common priors and signals are determined by the states of the world.

There are two states of the world, represented by two types of bingo cages in the experiment. Both types of cages contain 100 balls and act as possible random devices to generate private signals for all participants in the market. To avoid cognitive differences in private information, I use more neutral labels of signals — red ball and blue ball, corresponding to Y-type and N-type in model setup. At the beginning of the experiment, all participants observe the state space and how signals are generated for each possible state. They share a common prior $f(\omega)$, where ω in our experiment is the proportion of red ball in population. Once a bingo cage is chosen as a random device, the state of the world is realized. Then agents will receive a ball generated by the chosen cage. This procedure guarantees that signals are both "impersonal" and "informative." Agents who receive different balls expect different distributions of ω and agents who receive the same ball update ω the same way.

In the example of a typical Bayesian market illustrated in Figure 4.1, two typeA bingo cages contain 67 red balls, and two other cages contain 33 red balls. One cage will be randomly chosen to generate private signals, and therefore the common prior of the proportion of the red ball ω is:

$$f(\omega) = \begin{cases} \frac{1}{2} & \text{if } \omega = 0.67 \\ \frac{1}{2} & \text{if } \omega = 0.33 \\ 0 & \text{otherwise} \end{cases}$$



You are paired with 7 AAs

Round 1 of 14

00:31

1. One of four cages has been chosen for your group

3. Trading rules on the market

Liquidation Value is the number of "Red Ball" reports among other group members.

Market Price is a random number chosen between 0 to 7.

1. If you decide to send a "Red Ball" report and to be a **buyer**, a trade will occur if:

Your Buying Price >= Market Price>= MM's Selling Price

2. If you decide to send a "Blue Ball" report and to be a **seller**, a trade will occur if:

Your Selling price <= Market Price <= MM's Buying Price.

2. Draw a ball from the chosen bingo cage

Your draw from the chosen bingo cage is:

4. Buy or sell an asset on the market

I would like to send a "Blue Ball" report and to be a **Seller**

I would like to **Sell** an asset at price:

Figure 4.1: Decision screen

When a bingo cage is chosen (the state of the world is realized), an agent can click a button and draw a ball, showing in the left bottom area of the screen. Even though he does not know which cage has been chosen, he updates each state's likelihood and further infers the posterior probability of ω in the market. For instance, after receiving a red ball, a Bayesian agent expects a 67% chance of which the chosen bingo cage is a typeA cage.

$$f(\omega \mid t_i = Y) = \begin{cases} 0.67 & \text{if } \omega = 0.67 \\ 0.33 & \text{if } \omega = 0.33 \\ 0 & \text{otherwise} \end{cases}$$

Based on this posterior distribution, his expectation of ω is $E(\omega \mid t_i = Y) = 0.56$

Given the common prior and the private signal, each agent participates in a Bayesian market by submitting a report of the ball they receive and and a bid/ask price for an asset. The right bottom area of the interface shows their decisions. I allow for at most eight agents in the market and rescale the asset value, bid/ask prices, and the market price from $[0,1]$ to $[0,7]$. Bayesian markets also work for any $n \geq 4$.[8] The choice of $n = 8$ is a compromise between market interaction and experiment control. It generates multiple realizations of asset value but still controls the complexity of the task. The asset value is the number of red ball reports submitted by all opponent agents. The market price is randomly drawn from a truncated normal distribution on $(0,7)$. Such distributions increase the chance of transactions by weighting intermediate prices with higher probabilities.

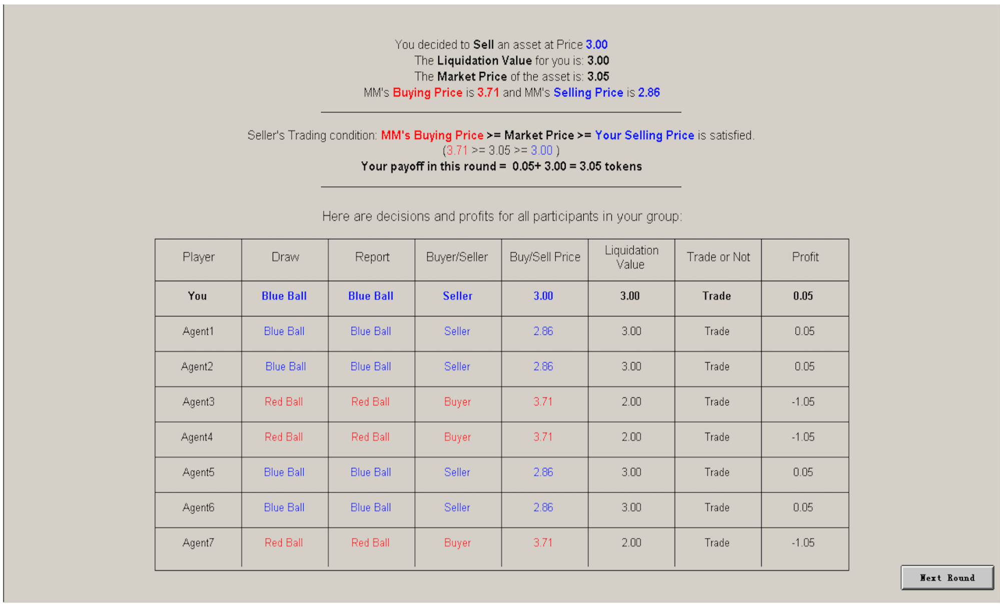
I design the experiment in a repeated setting. Each session consists of fourteen periods, and each period is a new Bayesian market. Appendix 4.A lists all sets of parameters for the priors and private signals. Although the Bayesian market is framed as a one-shot game, it is generally difficult for subjects to understand the game. They may not recognize the best response and the truth-telling equilibrium immediately but can do so in convergence. Learning in a repeated setting can be an effective de-biasing method. To facilitate learning, I keep the group members fixed during the session and provide each subject with the feedback regarding all players' signals, decisions, and profits after each period. Figure 4.2 is an example of the review screen.

It is important to emphasize that Bayesian markets reward truth-telling even when ground truth is not accessible. In our experiment, the truth, which is the ball generated by the chosen bingo cage, is verifiable. Potential relaxation of this procedure will be discussed later.

4.3.2 Experimental treatments

Given the truth-telling BNE of Bayesian markets, it is in a participant's best interest to report his private signal if he believes all other agents are truthful. However, there is no guarantee that subjects will hold such beliefs. For example, it's reasonable for agents to expect that some other agents are confused by the mechanism or prefer lying. This strict belief requirement of BNE might be the most arresting obstacle for truth-telling incentives. To understand whether and how Bayesian markets perform, I control subjects' beliefs over others' truthfulness in the experiment.

[8] One problem of a small sample is the possible manipulation of asset value through one's own report, which may encourage deception. Therefore, assets on Bayesian markets with the small sample are individually adjusted to ensure the incentive alignment of both signal and prediction reports. Specifically, each agent chooses to buy/sell an "individualized" asset whose value is the proportion of yes report among all other agents on the market. Also, there will be no trade for an individual when all other agents submit the same signal reports. With these adjustments, all theoretical predictions of Bayesian markets still hold: the BDM method incentivizes agents to submit truthful prediction reports; Y-type agents expect a higher asset value than N-type agents; truthful reports of singles for every agent is an equilibrium.



You decided to **Sell** an asset at Price **3.00**
The **Liquidation Value** for you is: **3.00**
The **Market Price** of the asset is: **3.05**
MM's **Buying Price** is **3.71** and MM's **Selling Price** is **2.86**

Seller's Trading condition: **MM's Buying Price >= Market Price >= Your Selling Price** is satisfied.
(3.71 >= 3.05 >= 3.00)
Your payoff in this round = 0.05+ 3.00 = 3.05 tokens

Here are decisions and profits for all participants in your group:

Player	Draw	Report	Buyer/Seller	Buy/Sell Price	Liquidation Value	Trade or Not	Profit
You	**Blue Ball**	**Blue Ball**	**Seller**	**3.00**	**3.00**	**Trade**	**0.05**
Agent1	Blue Ball	Blue Ball	Seller	2.86	3.00	Trade	0.05
Agent2	Blue Ball	Blue Ball	Seller	2.86	3.00	Trade	0.05
Agent3	Red Ball	Red Ball	Buyer	3.71	2.00	Trade	-1.05
Agent4	Red Ball	Red Ball	Buyer	3.71	2.00	Trade	-1.05
Agent5	Blue Ball	Blue Ball	Seller	2.86	3.00	Trade	0.05
Agent6	Blue Ball	Blue Ball	Seller	2.86	3.00	Trade	0.05
Agent7	Red Ball	Red Ball	Buyer	3.71	2.00	Trade	-1.05

Figure 4.2: Review screen

I construct the market with two types of agents — Human Agent (HA) and Algorithm Agent (AA). HAs are human participants. AAs are programmed to report private signals and posterior expectations of the asset value truthfully. To make participants believe that all (or some) other agents are truth-telling, I group HAs with AAs participating in Bayesian markets. Different numbers of HAs and AAs influence a human participant's belief of other agents' truthfulness. Following this simple idea, I design three treatments of Bayesian markets differing in the degree of the beliefs over others' truthfulness.

The first treatment, called 1HA treatment, aims to test whether Bayesian markets will induce the best response of truth-telling from individuals when their opponents are restricted to be truthful. Each human participant in the market groups with seven AAs. The task of an agent is to decide whether to buy or sell an asset and to submit a bid or ask price. Since AAs always tell the truth, human agents will reasonably believe that all other agents are truthful. Based on Prediction 4.2, it is optimal for a human participant to report his private signal and his posterior expectation of the asset value truthfully.

In the second treatment, called 3HA treatment, I allow for some disturbances of HAs' beliefs to test whether the best response of truth-telling from individuals is robust in Bayesian markets. Each human participant in the market groups with two other HAs and five AAs. AAs are still truthful, but HAs may lie. Given the small noises in the belief system, it is still optimal for a HA to report truthfully under strict participation condition[9] and if he expects a red ball type agent are more likely to report red ball than a blue ball type agent.[10] I design group composition and prior parameters to make sure that this condition is satisfied as much as possible. The simulation of 10000 draws shows that in 90% cases, it indeed is satisfied.

The third treatment, 8HA treatment, further tests whether Bayesian markets will select out the truth-telling equilibrium. Each human agent participates in a Bayesian market with seven other HAs. There are significant disturbances in participants' belief systems. They may form different beliefs over other HAs' truthfulness. However, they are rewarded for formulating the correct beliefs and further coordinating at the truth-telling BNE.

4.3.3 Experimental procedures

I ran the experiment in May and June 2016 at Erasmus University of Rotterdam. A total of 87 subjects were recruited in four sessions, and each session lasted around 90 minutes. The average payment was 21.80 euros. Table 4.1 summarizes the number of subjects, groups, and average payoffs for each treatment.

	1HA	3HA	8HA
Subjects	25	30	32
Groups	25	10	4
Observations	350	420	448
Payoffs	22.78	21.22	21.57

Table 4.1: Summary of three experimental treatments

Upon arrival, each subject was randomly assigned an ID and was guided to a computer desk. Then they were asked to read instructions[11] and finish a quiz regarding trade and profits in Bayesian markets. After that, all subjects would trade in markets for 14 periods. Each period is a new Bayesian market with different common priors and private information. At the end of the experiment, I also asked subjects to fill out a non-incentivized questionnaire regarding their understanding of the experiment, their socio-demographic characteristics, and self-reported risk attitudes.

The monetary unit in the experiment is called tokens, each worth 0.5 euro. After a market had closed in one period, asset price was randomly chosen from (0,7). Then a

[9] Strict participation means that an agent who expects strictly positive payoff will participate in the market. This condition is satisfied in truth-telling equilibrium.

[10] This result can be found in Baillon (2016).

[11] The instruction for 1HA treatment is attached in Appendix 4.B.

market maker calculated the average bid/ask price and determined whether a trade would occur for each agent. She would also liquidate all assets in the market and calculate participants' profits. Since agents might lose money, they were endowed with three tokens at the beginning of each period. The total payment for each subject was the sum of endowments and his profits in all 14 rounds. Subjects in 3HA and 8HA treatment faced similar interfaces and decision tasks. The only difference was the status bar showing different group compositions in the decision screen. To better understand how agents behave on Bayesian markets under different belief systems, I kept three treatments comparable in all aspects: parameter settings for common priors were the same; each set of parameters appeared in the same order; prices were determined by same distributions.

4.4 Experimental results

4.4.1 Aggregate truthfulness rate

The disturbances in people's belief systems do influence the validity of Bayesian markets. At the aggregate level, the proportion of truthful reports varies with the treatment stimulus. Table 4.2 summarizes the average percentages of truthful reports in three treatments. In 1HA treatment, 80% of the reports submitted by agents are their private signals. And this number is 68% for 3HA and 63% for 8HA Treatment, respectively.

	1HA	3HA	8HA
Truthfulness rate	0.80	0.68	0.63
Observations	350	420	448

Table 4.2: Aggregate rate of truthful reports in three treatments

Unsurprisingly, all three truthfulness rates are lower than the theoretical value of 100%. Even though the experiment satisfies prior and information assumptions of Bayesian markets, there are other implied structural assumptions imposed on agents for the prediction that truth-telling is a best response in 1HA and 3HA treatment and a BNE in 8HA treatment. For instance, subjects are required to use sophisticated Bayesian reasoning to predict others' signals in the same way. Given the noise inherent in real settings, the truthfulness rate in 1HA treatment provides reasonable supports for the validity of Bayesian markets.

Pairwise comparisons between aggregate truthfulness rates in different belief settings reveal valuable information about the performance of Bayesian markets. In 1HA treatment, 80% of the reports submitted by agents are the same as private signals. But this number is much lower in 3HA and 8HA treatment. Mann-Witney tests show that the truthfulness rate in 1HA treatment is significantly higher than that in 3HA and 8HA treatment, but there is no significant difference between 3HA and 8HA. These results imply that when there are more HAs in the market, participants feel uncertain about others' truthfulness, and Bayesian markets are less effective in inducing truth-telling.

4.4.2 Individual truthfulness rate

The treatment differences in truthfulness rates remain at the individual level. Figure 4.3 depicts the frequency of truth-telling for each subject in three treatments. In 1HA treatment, 25% of subjects report their private signals truthfully, and 21% only lie once during 14 periods. However, in 3HA and 8HA treatment, few subjects are fully truthful, and as a consequence, the aggregate truthfulness rates are lower than that in 1HA treatment. In terms of inducing truth-telling from individuals, again, I find Bayesian markets outperform in 1HA treatment.

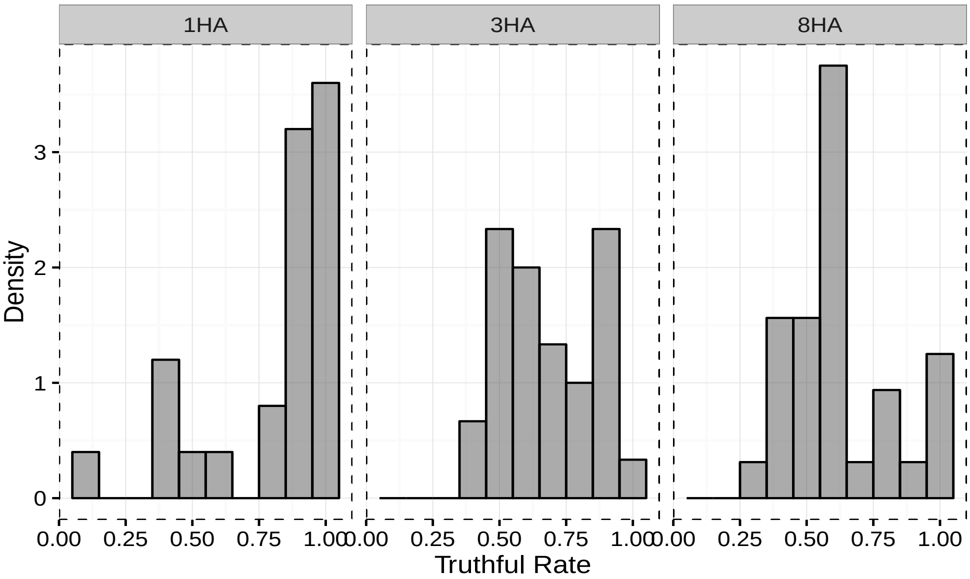


Figure 4.3: Histogram of individual truthfulness rate in three treatments

What causes the different truth-telling performances across treatments? Since the truthfulness rate captures the average level rather than a dynamic change of truth-telling, heterogeneity in strategy evolution among treatments may play a role. Another possible source for treatment effects is the heterogeneity in signal effects. It is defined as how truth-telling incentives vary with different signals. When the number of HAs increases, there will be more noise in the market, triggering different signal effects between treatments. The following subsections check these two possible sources.

4.4.3 Heterogeneity in strategy evolution

Report strategies in three treatments may follow different evolution patterns. For instance, truthfulness rates in three treatments may start at a similar level, but evolve at different speeds and therefore result in different aggregate levels. Since cognitive sophistication required by each experiment is increasing with the number of HAs in the market, I expected learning speed to be the highest in 1HA treatment, medium in 3HA and the lowest in 8HA treatment.

Figure 4.4 illustrates the time series of the average truthfulness rate in three treatments. The solid line is for 1HA treatment, dotted for 3HA, and dashed for 8HA treatment. There is little evolution trend in both 1HA and 8HA treatment. In 1HA treatment, the truthfulness rate is persistently high, limiting the room for further improvement. In 8HA treatment, the interaction among human agents introduces too much noise for effective learning. However, there is a prominent learning trend in 3HA treatment. The truthfulness rate starts at the lowest level of 57% and increases to 80%. Through learning, subjects in 3HA achieve a truth-telling rate close to that in 1HA treatment.

I do find heterogeneity in strategy evolution among treatments, but it is between 3HA and 1HA (8HA) treatment. Even though their truthfulness rates are similar, Bayesian markets with 3HA may outperform those with 8HA in the long run because of the learning effect. However, strategy evolution fails to explain treatment effects between 1HA and 8HA treatment.

4.4.4 Heterogeneity in signal effects

Another source of treatment effects is the heterogeneity in signal effects, meaning agents have different responses to different private signals. I calculate the posterior truthfulness rates conditional on private signals in Table 4.3. Given the private signal is a red ball, more than 80% of reports are truthful, and it holds for all three treatments. However, after receiving a blue ball signal, agents are more likely to lie. 40% of the reports submitted by agents in 3HA are untruthful, and this number increases to 49% in

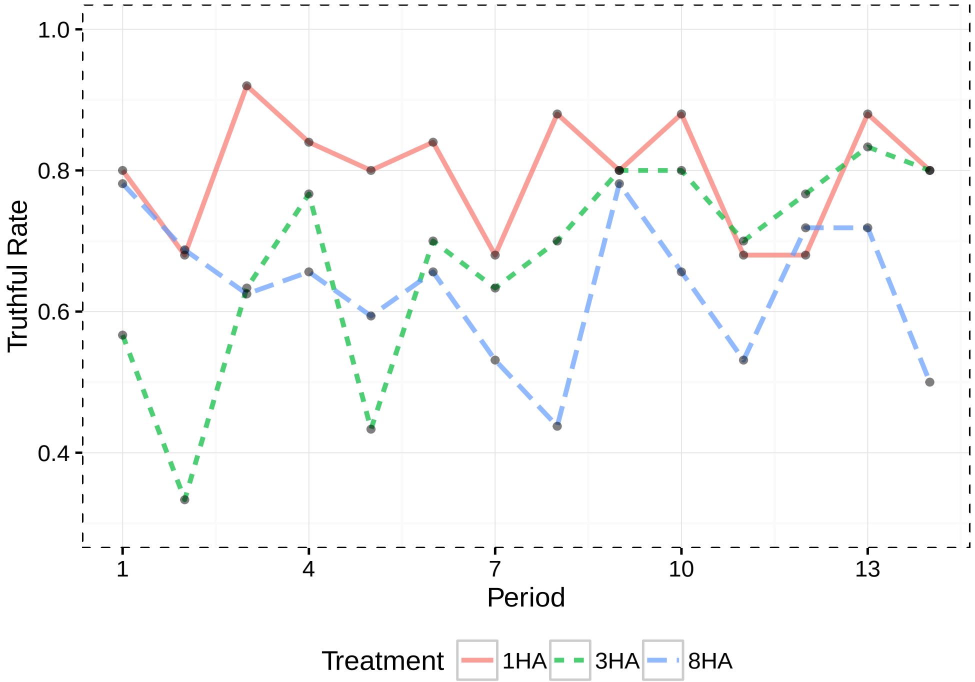


Figure 4.4: Truthfulness rate by periods

8HA treatment. Mann-Witney tests between two conditional truthfulness rates show significant differences in 3HA and 8HA treatment, but not in 1HA treatment.

	1HA	3HA	8HA
Pr(Truthful | Sig = Red)	0.85	0.81	0.85
Pr(Truthful | Sig = Blue)	0.77	0.60	0.51

Table 4.3: Truthfulness rate conditional on private signals

This result is quite puzzling because different labels of private signals should not affect truthfulness rates in a systematic manner. Moreover, I design the experiment with symmetric signals. First, it is equally likely to receive a red ball or a blue ball on average. Second, each market in the experiment corresponds to another market where labels of two signals are swapped. Therefore, the truthfulness rates should be equal for both types of signals.

However, the inherit elicitation procedure of Bayesian markets may lead to different perceptions for different signals. Specifically, signal reports submitted by agents are associated with their buy/sell decisions. Implied transaction positions of private signals, rather than labels, may affect incentives of truth-telling. Since these incentives are triggered by market conditions, they are highly likely to vary with group compositions for each treatment. A plausible explanation is that agents are more likely to buy than short sell assets because they are more experienced with purchase decisions. This phenomenon is quite normal in market institutions, and I call it "buying inclination". In the experiment, as the number of HAs increased, markets become noisier. The choice of buying/selling is more complicated, and thus there is a more severe buying inclination in 3HA and 8HA treatment.

I further test the buying inclination and its implication on truthfulness rates in Table 4.4. The benchmark of the buying rate is around 0.36. It is the realized frequency of receiving the signal of a red ball. If all participants' reports are truthful, 36% of them will buy an asset. However, buying rates calculated from agents' decisions in three treatments are all higher than their benchmarks. In particular, buying rates in 3HA and 8HA treatment are as high as 55% and 63%, implying a more severe bias towards buying than in 1HA treatment.

The third and fourth row of Table 4.4 report the truthfulness rates conditional on buyer/seller decisions. Among all reports from sellers, the majority of them are truthful. However, among all buyer reports in the market, there is a significant difference between 1HA and 3HA (8HA) treatment. 50% of the reports submitted by buyers are untruthful in 8HA treatment, compared with 53% in 3HA and 67% in 1HA treatment.

	1HA	3HA	8HA
Benchmark	0.36	0.36	0.37
Buying Rate	0.45	0.55	0.63
Pr(Truthful | Buyer)	0.67	0.53	0.50
Pr(Truthful | Seller)	0.90	0.85	0.86

Table 4.4: Buying rate and truthfulness rate

The aggregate truthfulness rate is an average of truthfulness rates of buyers and sellers with the buying rate as a weight. Taken together, I conclude that treatment effects are driven by signal effects. First, agents in 3HA and 8HA treatment are more likely to buy than sell assets; second, buyers are more likely to lie than sellers.

4.5 Explaining treatment effects: bid and ask prices

In this section, I will explain the differences in truthfulness rates in three treatments via agents' bid and ask prices for the asset in Bayesian markets.

4.5.1 Aggregate bid and ask prices

Figure 4.5 demonstrates the time series of the submitted price, the theoretical price, and the value of the asset in Bayesian markets. The solid line is the average price submitted by subjects in each period, representing an average HAs' ex-ante valuations of the asset. The dashed line shows the correct predictions based on their signals. And the dotted line is the ex-post asset value.

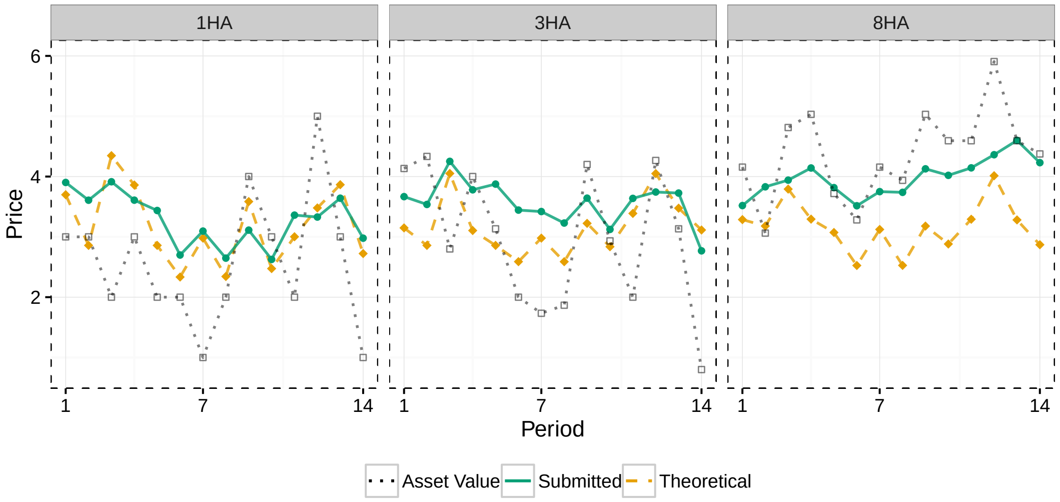


Figure 4.5: Aggregate bid/ask price and the theoretical price

I focus on the ex-ante and ex-post prediction gaps in each treatment. The first one means the difference between the submitted and the theoretical prices, and the second one is the difference between the submitted prices and the realizations of asset value. First, the ex-ante prediction gaps in all treatments are positive for all fourteen periods, meaning that an average HA in a Bayesian market predicts a higher asset value than a AA does. Therefore, HAs are more likely to buy assets than AAs in all treatments, consistent with the buying inclination shown in the previous section.

There is a significant difference between 1HA (3HA) and 8HA treatment in terms of the size and trend of the ex-ante prediction gaps. In both 1HA and 3HA treatments, ex-ante prediction gaps are quite small and exhibit no trend. In other words, an average HA in these two treatments correctly predicts the realization of asset value. While in 8HA treatment, the prediction gap starts at a substantial level and is increasing over time, indicating that HAs are persistently over-predicting asset value. There is also a significant difference between 1HA (3HA) and 8HA in their ex-post prediction gaps. However, the direction is opposite to the ex-ante gaps. An average HA in 8HA treatment predicts the asset value more accurately than the agent in 1HA and 3HA treatment does.

The treatment effects of both ex-ante and ex-post prediction gaps are consistent with each other and jointly explain the trend-chasing of submitted prices in 8HA treatment. Since subjects in 8HA treatment over-predicts asset value more severely, they are over-buying assets in the markets. This further drives up the ex-post realization of the asset value. In other words, their beliefs of the asset are self-confirmed by their decisions, leading to bubbles in Bayesian markets.

How does the trend of ex-ante and ex-post prediction gaps relate to the truthfulness rate? Figure 4.6 shows the same time series of price gaps for truthful and untruthful reports. It confirms that untruthful subjects are the main drive for the over-prediction of asset value and the resulting over-buying in markets. On average, truthful agents submit prices close to theoretical ones. They act as Bayesian agents who exploit private information to update belief. Untruthful agents, on the other hand, severely over-predicts asset value. They act as trend-chasers who take into account possible bubbles in the market. For 8HA treatment, comparisons between truthful and untruthful agents are more evident: truthful subjects submit predictions reflecting the fundamental value of the asset and untruthful subjects chase bubbles and submit predictions reflecting the market valuation of the asset.

Since untruthful reports are in the form of either receiving a red ball but reporting blue ("RB") or receiving a blue ball but reporting red ("BR"). I further check the time series of the submitted price, theoretical prices, and asset value based on private signals and buy/sell decisions in 8HA treatment in Figure 4.9 of Appendix 4.C. The patterns are similar. Agents who receive blue ball signals and decide to be buyers are trend-chasers. They submit prices close to the realization of asset values. Agents who receive red ball signals and choose to be sellers act as Bayesians. Their prices reflect that they formulate correct beliefs in asset value after knowing their private information.

Combining with the buying inclination in the previous section, I find that agents in 8HA treatment who receive blue ball signals have higher valuations of the asset than its fundamental value. They thus are more likely to buy assets, further raising the realized asset value. The asset value indeed confirms their belief, and they would continue to chase the trend and finally cause bubbles in the market.

Another point is that the time series of prices are aggregated with four Bayesian markets, each corresponding to a group of 8HAs in the treatment. Appendix 4.C shows the same time series of submitted price, theoretical price, and asset value for each group. The aggregate ex-ante prediction gaps are mainly driven by Group 24. In all three

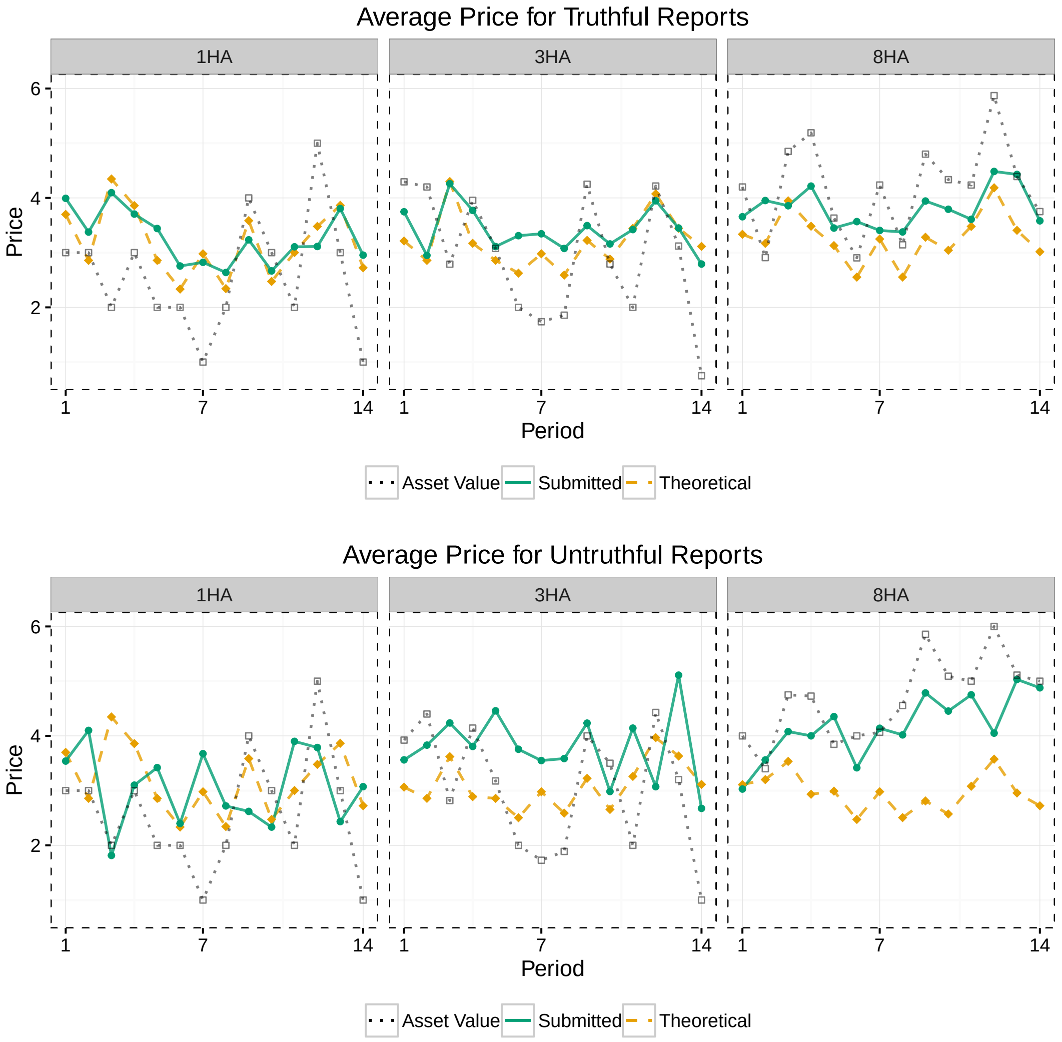


Figure 4.6: Aggregate bid/ask price and the theoretical price by truthfulness

markets, there is no clear evidence of trend-chasing and the resulting market bubbles. Deleting Group 24 will reduce the extend of bubbles in 8HA treatment.

4.5.2 Shirking, speculation, and updating bias

A Bayesian agent will update his private signal and then submits his expected number of red reports as a bid (or ask) price. When he believes that all others are truthful, his bid/ask price is the expected proportion of red balls among seven agents $7E(\omega \mid t_i)$ (Prediction 4.1). In the experiment, ω takes two values, ω_A and ω_B. Each one represents the proportion of red balls in the corresponding types of the cage. Therefore, the bid/ask price should be bounded by two values of 7ω. Moreover, subjects' posterior beliefs about the state A (the chosen cage is of type A), denoted as $\pi(A \mid t_i)$ can be backed out by their bid/ask prices through $c_i = \pi(A \mid t_i)\omega_A + (1 - \pi(A \mid t_i))\omega_B$.

The bid/ask prices reveal agents' updating and participation decisions in a Bayesian market. Different patterns may explain how belief environment affects the validity of Bayesian markets. For instance, if an agent submits a bid lower than $7\omega_B$ or an ask higher than $7\omega_A$, he cannot trade an asset whatever the asset price is, and others' decisions are. In other words, agents in a Bayesian market can avoid trading an asset by submitting low bids or high asks on purpose. I call such reports being "shirking". On the other hand, If an agent submits a bid higher than $7\omega_A$ or an ask lower than $7\omega_B$, he will guarantee himself a trade of the asset. I call such reports being "speculative". When the bid/ask price is within the range of $7\omega_B$ and $7\omega_A$, the agent's posterior belief of the state is rationalized by a regular probability on $[0,1]$, I call such reports being "Bayesian".

Table 4.5 reports the proportion of each types of bid/ask patterns. Around 10% of bid/ask prices show that agents prefer no participation in the market. Agents in 3HA and 8HA treatment are significantly more likely to be speculative. In 8HA treatment, most of the speculative reports are buyers who submit very low bid prices to make sure successful trades. This is consistent with the trend-chasers in the market. The majority of subjects valuations of the asset are still justified by Bayesian reasoning, and agents in 1HA treatment are significantly more Bayesian.

	1HA	3HA	8HA
#Obs	350	420	448
Shirking	9%	10%	9%
Speculative	15%	22%**	29%**
Bayesian	75%	67%**	62%**
#Bayesian	265	283	279

Note: **p<0.05;

Table 4.5: Proportion of different types of bid/ask price patterns.

Does the difference in speculative and Bayesian patterns between 1HA and 3HA (8HA) treatment explain the treatment effects in truthfulness rate. After excluding shirking agents, the treatment effects remain. After further excluding speculative agents, the size of the treatment effects is smaller. The result implies that the disturbances in people's beliefs over others' truthfulness induce updating biases. Moreover, the speculative agents in the market further enlarge the biases. The interaction between updating biases and speculation motives raises the bubbles in the market and impedes the truth-telling in Bayesian markets.

I focus on the bid/ask prices in the category of "Bayesian" and further examine people's updating biases in three treatments. Are the implied posteriors consistent with Bayesian updating? If not, do they support lying in the transaction of an asset? According to Bayes' Theorem of belief updating, we have the following formula of the posterior odds after receiving a signal of red ball:

$$\frac{\pi(A \mid Red)}{\pi(B \mid Red)} = \frac{p(A)}{p(B)} \times \frac{p(Red \mid A)}{p(Red \mid B)}.$$

Taking logs to both sides of the equality, I estimate the updating biases associated with the prior information and the diagnostic information of private signals separately by the following Grether (1980) regression:

$$\ln \frac{\pi_{ij}(A \mid Red)}{\pi_{ij}(B \mid Red)} = \beta_0 + \beta_1 \ln \frac{p_j(A)}{p_j(B)} + \beta_2 \ln \frac{p_j(Red \mid A)}{p_j(Red \mid B)} + \varepsilon_{ij}.$$

The dependent variable is subject i's posterior odds after receiving a red ball in the period j. The independent variables include the prior odds and the likelihood ratio for the same problem j. When the subject is a perfect Bayesian, the coefficients for both prior odds and likelihood ratios are equal to one. Following the literature (see Benjamin (2019)), there are four types of updating bias. The coefficient of prior odds determines whether the subject has a bias of base-rate neglect ($\hat{\beta}_1 < 1$) or confirmation bias ($\hat{\beta}_1 > 1$). The coefficient of likelihood ratio determines whether the subject has under-inference ($\hat{\beta}_2 < 1$) or over-inference ($\hat{\beta}_2 > 1$) bias.

Table 4.6 reports the regression results. The coefficients of the log of prior odds are close to those in the literature. Subjects generally exhibit base-rate neglect. A joint F-test of $\hat{\beta}_1$ shows that the updating bias of base-rate neglect is similar across three treatments. The coefficients of the log of the likelihood ratio (diagnostic information) show that agents under-infer from the private signal. In 1HA treatment, the magnitude of the under-inference (0.77) is significant and comparable to those estimated in standard Bayesian tasks. However, estimates of $\hat{\beta}_2$ are not significant, indicating that agents in 3HA and 8HA treatment ignore the private information. A joint F-test of $\hat{\beta}_2$ shows that the under-inference biases are significantly different between 1HA and 3HA (8HA) treatment. The difference in updating biases is consistent with the prediction gaps in Bayesian markets. When there are speculative trends, agents chase the trend and ignore their private signals. In 3HA treatment, AAs will stabilize the trend, while in 8HA treatment, the trend persists and evolves market bubbles.

	Dependent: LogPostOdd		
	1HA	3HA	8HA
Constant ($\hat{\beta}_0$)	-0.08	-0.09	0.41^{***}
	(0.13)	(0.12)	(0.14)
LogPriorOdd ($\hat{\beta}_1$)	0.33^{*}	0.53^{***}	0.37^{**}
	(0.19)	(0.09)	(0.15)
LogDiagOdd ($\hat{\beta}_2$)	0.77^{**}	0.13	0.15
	(0.38)	(0.09)	(0.12)
Observations	265	283	279
Note:	*p<0.1; **p<0.05; ***p<0.01		

Table 4.6: Estimating the belief updating biases for three treatments

4.5.3 Heuristics

Keep and switch heuristics

A simple heuristic of learning is keeping or switching strategies conditional on profits in the previous period. Under this heuristic, subjects keep the strategy if it yields positive profit in the last round, and they switch to alternative strategies if otherwise. Figure 4.7 captures the frequencies of the switch and keep patterns for each treatment. There are four types of switch and keep pattern: "Keep truth-telling", "Keep lying," "Truth-telling to lying," and "Lying to truth-telling," each corresponding to the dotted, the solid, the dot-dashed and the dashed line. The truth-telling is quite focal in all treatments. The frequency of "Keep truth-telling" stays steady around 70% in 1HA treatment but is more volatile in 3HA and 8HA treatment. Remarkably, there is an increasing trend in 3HA for the pattern of "Keep truth-telling", which explains the learning effect of truthfulness rate in 3HA treatment. All the other three patterns are much volatile.

The change of frequencies of switch patterns may rely on previous profits. Table 4.7 lists the conditional switch rates for each treatment. I find more switches in both 3HA treatment and 8HA treatment. However, conditional on previous profits, there is no significant difference between switch frequencies. Collectively, when there are more noises in others' truthfulness, subjects try alternative strategies, switching from truth-

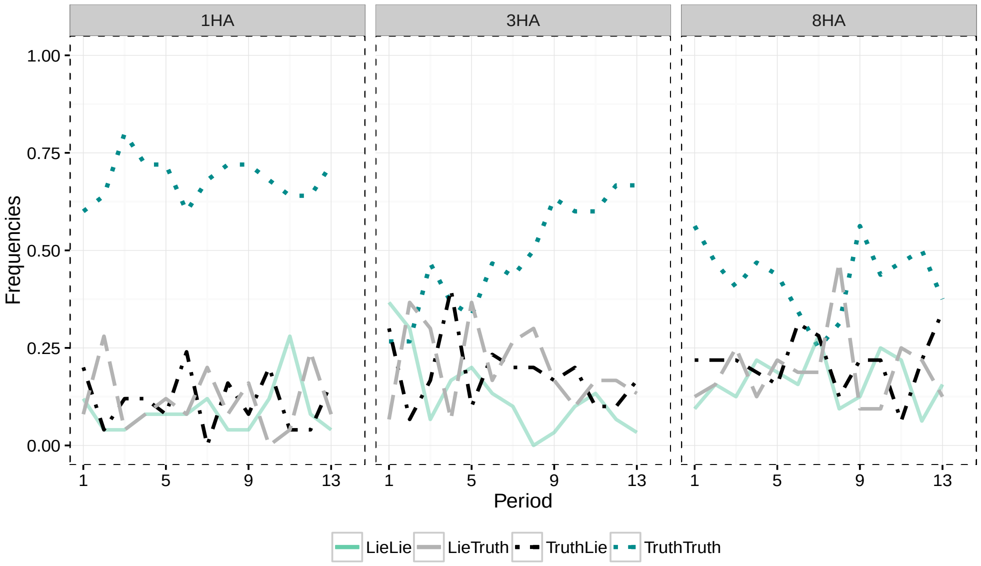


Figure 4.7: Time series of strategy keep and switch

telling to lying or from lying to truth-telling. However, these switches do not depend on previous profit.

Previous profit	Switch strategy	1HA	3HA	8HA
Positive	truth to lie	0.12	0.19	0.16
	lie to truth	0.04	0.21	0.28
Negative	truth to lie	0.11	0.19	0.25
	lie to truth	0.16	0.21	0.15
No Trade	truth to lie	0.11	0.18	0.22
	lie to truth	0.15	0.20	0.18

Table 4.7: Switch patterns conditional on previous profit

Imitation heuristic

Another simple heuristic is to imitate strategies that yielded the highest profit in the previous rounds. In the experiment, when each period completed, subjects learned all information about the last period on a review screen. They may check strategies and the corresponding profits of others and further adjust his strategy in the next round. Table 4.8 reports the imitation rate and winner's truthfulness rate for each treatment. Imitation rate is calculated as the frequency of which a subject's current strategy is the same as the winner's strategy in the previous round.

Treatment	1HA	3HA	8HA
Imitation rate	0.80	0.67	0.55
Winner's truthfulness rate	0.99	0.96	0.69

Table 4.8: Imitation rates in three treatments

First, winner's truthfulness rate are close to one in 1HA and 3HA treatment, implying truth-telling yielded the highest profit on average. This result is not surprising because of the existence of AAs in the market. AAs are always truthful and will be winners in the equilibrium. When there are no AAs to stabilize the market price, bubbles arise, and lying can be the winning strategy. Hence the truthfulness rate for winners in 8HA treatment is significantly lower.

Second, the imitation rate in 1HA treatment is 80%, significantly higher than that in 3HA and 8HA treatment. Notice that the imitation rate here captures coincidence, rather than causality between subjects' strategy and winners' strategy. Since it is impossible to isolate imitation from other heuristic or strategic considerations solely from revealed decisions, I cannot conclude that subjects in 1HA are more likely to imitate the winners' strategy. But the imitate rates show a higher correlation between subjects' strategy and the winner's strategy in 1HA treatment.

How does the imitation heuristic affect truth-telling in each treatment? Figure 4.8 depicts the scatter plot of the imitation rate and the truthfulness rate. The correlation is quite strong in 1HA and 3HA treatment. But in 8HA treatment, a high imitation rate may not imply (or be implied by) high truthfulness rate because the winning strategy is less likely to be truthful.

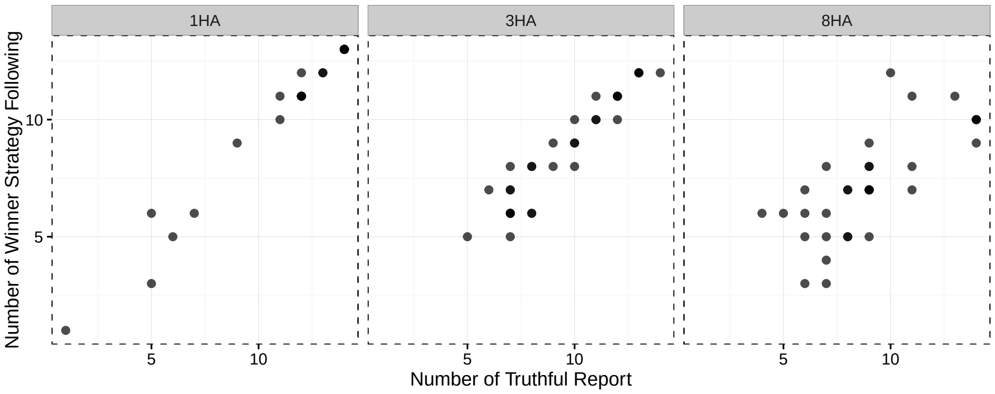


Figure 4.8: Relationship between imitation rate and truthfulness rate

4.6 Conclusion and discussions

This paper tests the validity of Bayesian markets. Will they induce truth-telling, either as a best response or as equilibrium from individuals? I construct Bayesian markets in an experiment and manipulate human agents' beliefs in other players' truthfulness in three settings. I find Bayesian markets effectively induce truth-telling as a best response when agents reasonably believe that all their opponents are also truthful. In particular, agents submit truthful reports of private signals on a large scale (80%) and form correct posterior expectations of the asset value in the market. However, when agents suspect that some of their opponents may lie, Bayesian markets are less effective. Participants in 3HA and 8HA treatment are less likely to report their private signals, and they are more likely to over-predict the asset value.

The appearance of bubbles in 3HA and 8HA treatments explains how the belief noises affect the performance of Bayesian markets. Even though the majority of subjects' posterior beliefs after knowing their private signal are rationalizable by Bayesian reasoning, they reveal more severe under-inference when there are noises in their belief systems. Subjects in 3HA and 8HA treatment ignore their private signals and focus on the trends of asset realizations. I interpret the process of bubbles as follows: (1) Common priors and impersonal private information are not sufficient to induce common posterior expectations. (2) Due to the speculative buyers in the market, traders who are supposed to sell assets predict higher-than-fundamental asset value and thus are more likely to buy assets. (3) over-buying in the market raised the realization value of assets, which further confirmed ex-ante belief. (4) more traders will ignore the private signal and chase the trend to buy assets. AAs in 3HA treatment manage to stabilize the speculative trend. While in 8HA treatment, bubbles raise in the Bayesian markets.

A closer check of what generated the bubble in the first place revealed buying inclination among participants, which also explained a relatively lower frequency of truth-telling in 3HA and 8HA treatment. To be specific, when traders are more uncertain about the behaviors of other traders, they would be more likely to buy than short sell assets simply because of the familiarity of the former context. The speculative buyers reflected in the bid/ask patterns also confirm the buying inclination. Initially, it might just raise the average expectation of asset value slightly, but it had the potential to trigger self-confirming belief in the market and to bring about bubbles. A critical lesson for the practical implementation of the Bayesian market mechanism is that I may improve data quality by familiarizing participants with the concept of short-sell.

One concern for the validity of Bayesian markets is the seemingly forced participation in our experiments. In the truth-telling equilibrium, Bayesian markets predict that all agents, regardless of their signals, choose to participate in the market since truth-telling yields a strictly positive payoff in expectation. However, when an agent believes that the market is out of equilibrium, he may prefer to opt-out. For instance, even though an agent with a blue ball signal recognizes that there are bubbles on the market, he is forced to "ride the bubble" because otherwise, he may lose. It should be noted that traders can opt-out in our experiment by submitting a bid of 0 or an ask of 7. However, few subjects did so (2.5% for buyers and 1.8% for sellers). The analyses of bid/ask patterns show that the out-out rates are not the drive for the difference in truthfulness rate in three treatments. Many subjects may not realize there is an option to step out of the market. By introducing treatment with a salient button for opt-out, I may increase the validity of Bayesian markets by mitigating bubbles in the market.

Even though Bayesian markets show promises in inducing truth-telling as a best response under the perfect belief, there is no benchmark for its validity in our experiment. Considering the requirement of sophisticated Bayesian reasoning in Bayesian markets, respondents may not respond to financial incentives by telling the truth. Even if they do, the improvement in truth-telling may not be enough to justify the application of Bayesian markets in practice. Therefore, I may better evaluate the validity of Bayesian markets with the help of a benchmark treatment, where subjects receive fixed payment in each period.

How well Bayesian markets perform relative to other truth-telling mechanisms is another related question. A natural candidate for comparison is BTS. In the same setup of priors and private information, it is possible to design comparable experiments of BTS. Instead of buying/selling assets, subjects are incentivized by information scores in BTS. Another possible candidate is the peer prediction method, which is possible due to the design of states and priors in our experiment, where mechanism implementers can learn common priors, even though it is not essential for Bayesian markets.

An important direction for future work is to test Bayesian markets with subjective truth. The experimental design is very challenging. On the one hand, tasks should be subjective enough for agents to believe that experimenters are impossible to verify the underlying truth. On the other hand, they should not be too subjective to be evaluated. One potential approach is to engage participants in tasks with partially subjective truth. For instance, it contains verifiable characteristics such as the range, mean, or distribution.

Appendix 4.A Parameter settings for the experiment

	C_A	ω_A	ω_B	ω	$E(\omega \mid R)$	$E(\omega \mid B)$	AAbid	AAask
1	0.50	0.70	0.30	0.50	0.58	0.42	0.55	0.39
2	0.25	0.85	0.38	0.50	0.58	0.42	0.55	0.39
3	0.75	0.62	0.15	0.50	0.58	0.42	0.56	0.39
4	0.50	0.67	0.33	0.50	0.56	0.44	0.53	0.41
5	0.25	0.80	0.40	0.50	0.56	0.44	0.53	0.41
6	0.75	0.60	0.20	0.50	0.56	0.44	0.53	0.41
7	0.50	0.57	0.23	0.40	0.47	0.35	0.44	0.33
8	0.25	0.70	0.30	0.40	0.48	0.35	0.44	0.33
9	0.50	0.77	0.43	0.60	0.65	0.53	0.62	0.50
10	0.75	0.70	0.30	0.60	0.65	0.52	0.62	0.49
11	0.50	0.65	0.25	0.45	0.54	0.38	0.51	0.35
12	0.25	0.80	0.34	0.46	0.54	0.38	0.51	0.36
13	0.50	0.75	0.35	0.55	0.62	0.46	0.59	0.43
14	0.75	0.66	0.20	0.55	0.62	0.46	0.59	0.43

Table 4.9: Parameters for priors, information structures, and posteriors of AAs in fourteen periods

- C_A is the probability that the chosen cage for the group is of Type A.
- ω_A (ω_B) is the proportion of red ball in a cage of Type A (B).
- ω is the prior of the proportion of red ball signals among population.
- $E(\omega \mid R)$ ($E(\omega \mid B)$) is the posterior of the red ball proportion among population conditional on receiving a Red (Blue) Ball signal.
- AAbid (AAask) is the bid (ask) price submitted by AAs, which is adjusted from $E(\omega \mid R)$ ($E(\omega \mid B)$) due to small sample restriction in the experiment.

Appendix 4.B Experimental instructions (1HA)

Preliminary Remarks

Your payoff in today's experiment is directly linked to your decisions. Please pay careful attention to this instruction because it can help you to better understand the experiment.

During the experiment, please do not communicate with any other participants and do not look at their computer monitors. If you have questions at any time, please raise your hand and we will come to you promptly.

Overview

The currency in the experiment is token and the conversion rate between tokens and euros is: 1 euro = 4 tokens. The experiment consists of 20 rounds. In each round, you will be endowed with 4 tokens to participate in a market to trade an asset. Your tasks are to: (1) decide whether to buy or sell an asset; (2) provide prices at which you would like to buy or sell the asset. These two decisions jointly influence the possibility of whether a trade will occur and the associated profits. Your payoff in each round is the sum of the endowment and your profit. The total payoff in the experiment is the sum of payoffs for the 20 rounds.

During the experiment, you will be grouped with seven Algorithm Agents (AAs). They are programmed identically and will face the same tasks as you do.

There is a "Market Maker" (MM) on the market, who collects all agents' decisions in your group and formulates his buying/selling price. You can trade at most one asset with MM.

At the end of each round, the asset will be liquidated[12]. If there is no trade, your profit is 0. If a trade occurs, a buyer's profit is the difference between the *Liquidation Value* and the *Market Price* and a seller's profit is the difference between the *Market Price* and the *Liquidation Value* .

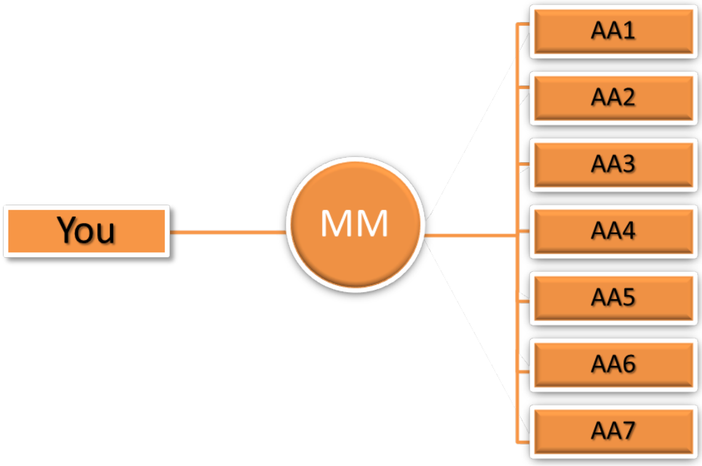
At the end of the experiment, you will be paid privately. When we call your ID number, please come forward to the sign-in counter and receive your earnings. We ensure you that any information regarding your participation, your name, ID number and payoffs are kept strictly confidential.

[12] Asset Liquidation (or settlement), is the delivery of an asset by a buyer or a seller.

How does the Market Work?

What is the Structure of the Market?

The structure of the market is shown in the following figure. You do not directly trade with other agents in your group, but with MM. That is, if you choose to be a buyer, MM will act as a seller; similarly, if you decide to be a seller, MM will act as a buyer. However, decisions of other members in your group will influence your trading probability, the *Liquidation Value* and your profits.

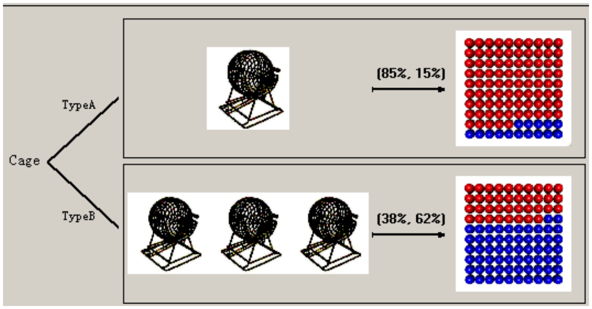


How to Make Decisions?

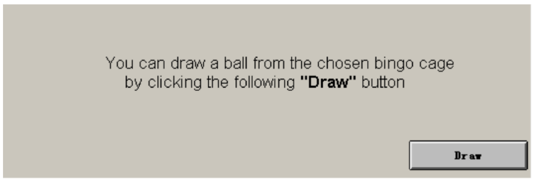
Your decision processes are described in the following steps:

1. **One of four bingo cages has been chosen for your group**

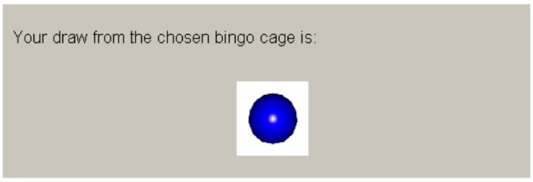
   Prior to the start of each round, the computer has already randomly chosen *one* from *four* bingo cages for your group. All of them contain 100 balls of red and blue, but they differ in the proportions of the two colors of balls. For example, in the following figure, a cage of Type A contains 85 red balls and 15 blue balls; three other cages are of Type B and contain 38 red and 62 blue balls. The chosen bingo cage determines the **SAME** probability with which you and your opponents draw a red or a blue ball.



2. **Draw a ball from the chosen bingo cage**
   Without knowing the type of the chosen bingo cage, each agent in your group will draw a ball from it **with replacement**. That means, **every agent in your group has the same probability of drawing a red or a blue ball**. In the following figure, a Blue Ball is drawn from the chosen cage. You will not know any other agent's draw and similarly, your draw will not be revealed to any other agent. However, from your own draw, you can infer which bingo cage could be the chosen one and what are possible draws for your opponents.

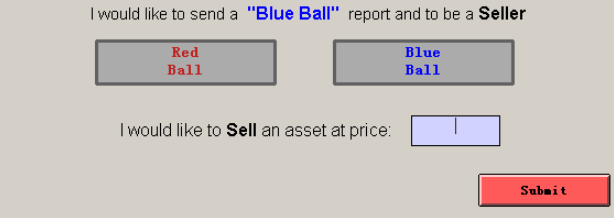


You can draw a ball from the chosen bingo cage by clicking the following **"Draw"** button



Your draw from the chosen bingo cage is:

3. **Buy or sell an asset on the market**
   After the draw, you need to choose to either buy an asset by sending a "Red Ball" report or sell an asset by sending a "Blue Ball" report. You also need to specify the buying and selling price of the asset, as in the following figure:



I would like to send a **"Blue Ball"** report and to be a **Seller**

I would like to **Sell** an asset at price:

What is the Asset on the Market?

The *Market Price* of the asset is randomly drawn from the interval (0, 7) and will be revealed at the end of each round.

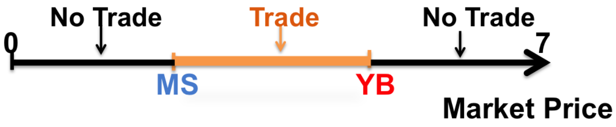
The *Liquidation Value* of the asset is determined by your opponents' reports. It is equal to the number of "Red Ball" reports among your 7 opponents.

When does a Trade Occur?

1. **Buyer's Trading Condition**
   If you decide to send a "Red Ball" report, you will automatically become a buyer and you can successfully buy an asset from MM at the *Market Price* if:

   $$\boxed{\textbf{Your Buying Price} \geq \textbf{Market Price} \geq \textbf{MM's Selling Price}},$$

   where **MM's Selling Price** is the average selling price among your opponents.

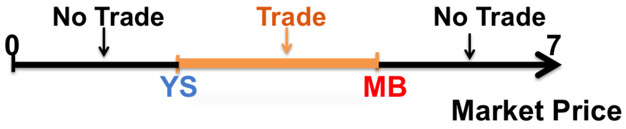
   The above figure shows a buyer's trading condition for any possible *Market Price* on (0,7). Suppose MS is MM's Selling Price and YB is Your Buying Price, only if the realization of the *Market Price* is between MS and YB, there will be a trade. The associated profit for you is:

   $$\boxed{\textit{Buyer's Profit} = \textit{Liquidation Value} - \textit{Market Price}}$$

2. **Seller's Trading Condition**
   If you decide to send a "Blue Ball" report, you will automatically become a seller and you can successfully sell an asset to MM at the *Market Price* if:

   $$\boxed{\textbf{Your Selling Price} \leq \textbf{Market Price} \leq \textbf{MM's Buying Price}},$$

   where **MM's Buying Price** is the average buying price among your opponents.

   The above figure shows a seller's trading condition. YS is Your Selling Price and MB is MM's Buying Price. There will be a trade for you only if the realization of the *Market Price* falls between YS and MB. The associated profit is:

   $$\boxed{\textit{Seller's Profit} = \textit{Market Price} - \textit{Liquidation Value}}$$

3. **Extra Condition**
   When all your opponents choose to buy the asset, No Trade will occur for you. Similarly, No Trade condition applies when all your opponents choose to sell. If there is No Trade, the associated profit is:

$$\boxed{\textit{Profit for No Trade} \;=\; \mathit{0}}$$

How do Algorithm Agents Make Decisions?

The Algorithm Agents (AAs) are programmed to follow certain decision rules. First, after he draws a ball, he will report the same color of ball as his draw. If he draws a ”Red Ball”, he will be a buyer by sending a ”Red Ball” report, and if he draws a ”Blue Ball”, he will be a seller by sending a ”Red Ball” report. Second, given his draw, he will calculate an expectation of the number of agents who drew a red ball and report it as his buying/selling price.

An Example

Suppose your opponents' decisions are shown in the following table:

Table 4.10: Opponents' Decisions

Agent	Draw	Report	Buyer/Seller	Buy/Sell Price
Agent 1	Red Ball	Red Ball	Buyer	4.20
Agent 2	Blue Ball	Blue Ball	Seller	3.50
Agent 3	Blue Ball	Blue Ball	Seller	3.50
Agent 4	Red Ball	Red Ball	Buyer	4.20
Agent 5	Red Ball	Red Ball	Buyer	4.20
Agent 6	Blue Ball	Blue Ball	Seller	3.50
Agent 7	Red Ball	Red Ball	Buyer	4.20

Liquidation Value = number of ”Red Ball” reports among your opponents = 4.00.
MM's Buying Price = average buying price among your opponents = 4.20.
MM's Selling Price = average selling price among your opponents = 3.50.

Suppose you want to buy an asset with an offer of 5.00 and the realization of the *Market Price* is 3.80, you can successfully buy an asset because $5.00 \geq 3.80 \geq 3.50$. You need to pay 3.80 and can liquidate this asset at 4.00, thus your profit is $4.00 - 3.80 = 0.20$ and your payoff is 4.20.

Suppose you want to sell an asset at price 3.90 and the realization of the *Market Price* is still 3.80, there will be no trade for you because $3.90 \geq 3.80 \leq 4.20$. If you lower your selling price to 3.30, you will successfully sell an asset to MM. Your profit in this case is $3.80 - 4.00 = -0.20$ and your payoff is 3.80.

Appendix 4.C Additional analyses on the bid/ask prices in 8HA treatment

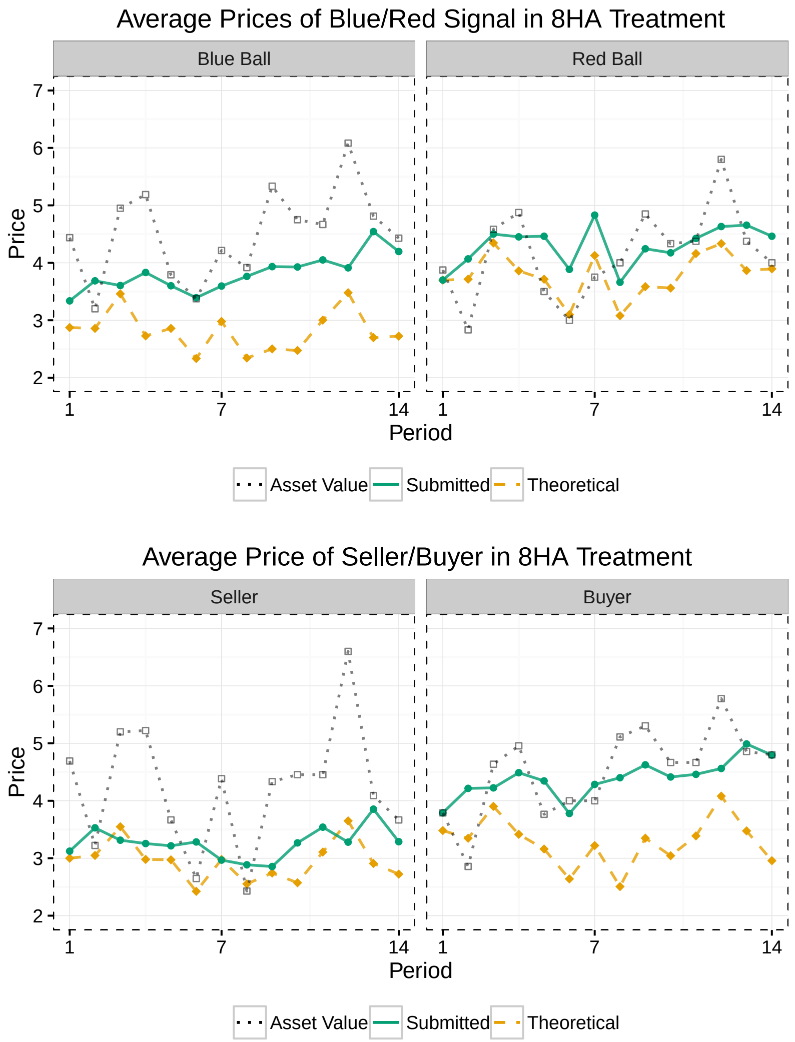


Figure 4.9: Price divergence in 8HA treatment by signals and buyer/seller decisions

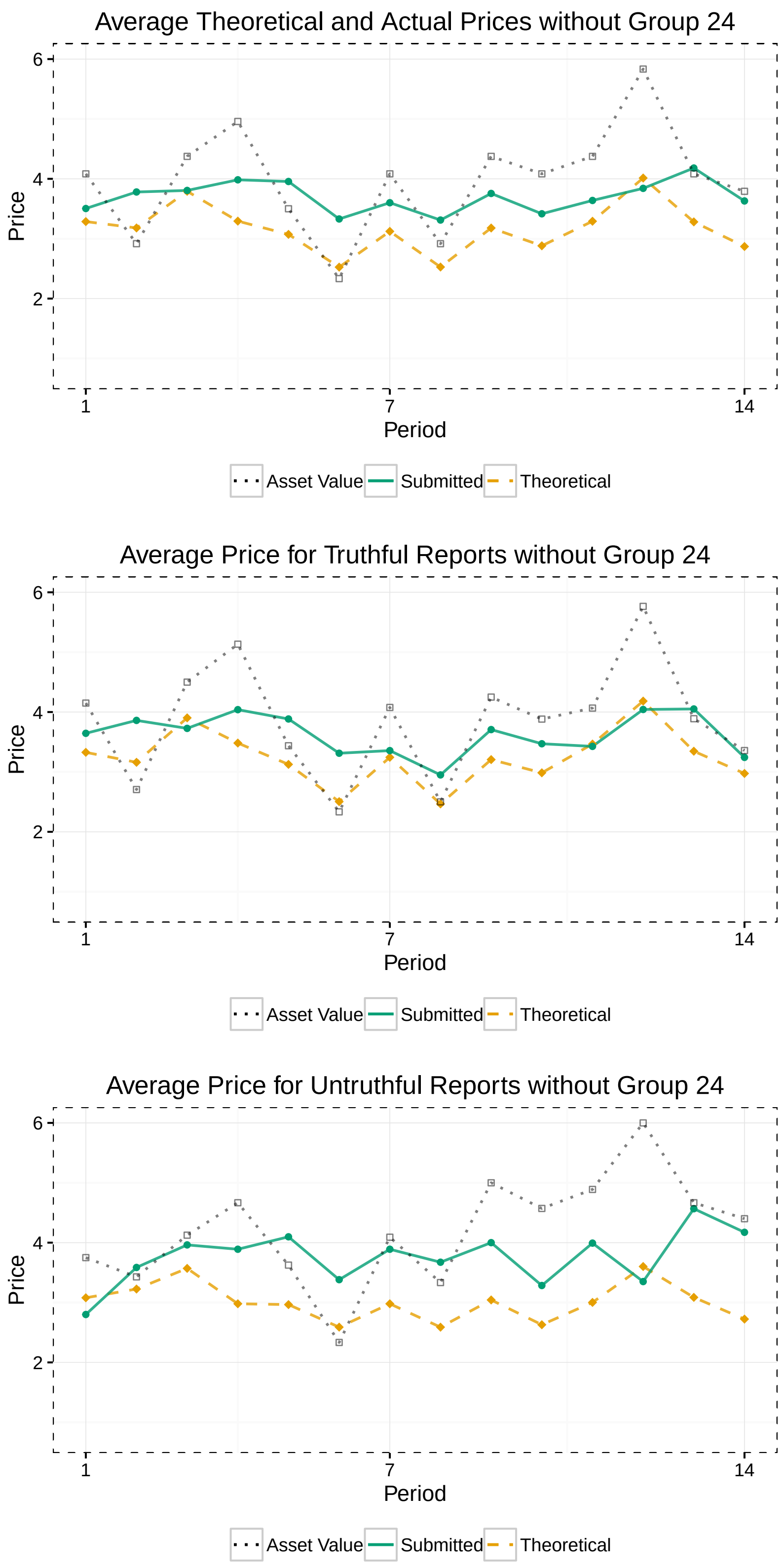


Figure 4.10: Aggregate prices without Group 24

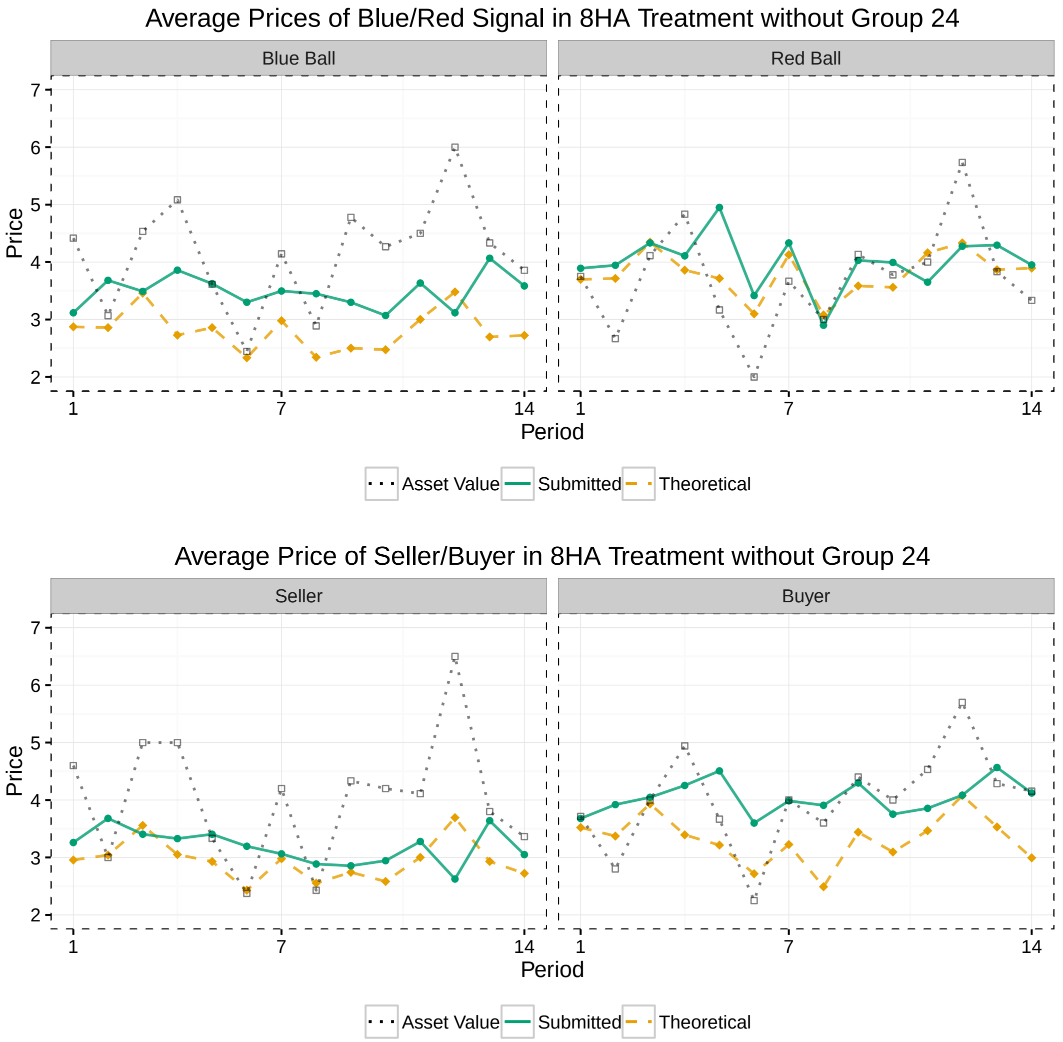


Figure 4.11: Average price by signals and buyer/seller without Group 24

Chapter 5

Simple bets to elicit private signals[1]

Abstract: This paper introduces two simple betting mechanisms, Top-Flop and Threshold betting, to elicit unverifiable information from crowds. Agents are offered bets on the ratings of an item about which they received a private signal versus that of a random item. We characterize conditions for the chosen bet to reveal the agents' private signal even if the underlying ratings are biased. We further provide micro-economic foundations of the ratings, which are endogenously determined by the actions of other agents in a game setting. Our mechanisms relax standard assumptions of the literature such as common prior, and homogeneous and risk neutral agents.

5.1 Introduction

Suppose the manager of a customer-care call center wants to assess her employees through some customer satisfaction measures. At the end of each call, she invites customers to take a one-question survey about whether or not they are satisfied with the services. She can reward participation with a small prize (voucher or fidelity points) but this is not enough. She would also like to have the customers think carefully about the question and provide truthful answers. If she were able to verify the answer, incentivizing truth-telling would be easy. However, only the customers themselves know whether they are actually satisfied or not, making it difficult to align rewards with truth-telling. We propose the following solution. The manager can reformulate the survey question and ask customers to bet whether the employee they talked to has a higher or lower satisfaction rate than another, randomly selected, employee from the call center. Customers who win the bet receive the prize.

[1] This chapter is based on Baillon and Xu (2019).

We call the aforementioned method Top-Flop betting and show that it provides incentives for agents to truthfully reveal private information. We consider two cases. In the first case, the bets are defined on a pre-existing satisfaction rating, which may be biased as long as it is informative enough (as specified later). In the second case, the rating is a function of the bets chosen by other customers. Another method introduced in this paper and that we call Threshold betting, induces truth-telling by making customers bet on which employee (the one they talked to or a random one) is more likely to get a satisfaction rate exceeding a given threshold.

It is easy to implement Top-Flop and Threshold betting in many settings in which people receive private binary signals, in the form of tastes or experiences. An application, which we will use as a leading example, is to elicit whether people liked or disliked a movie after previewing it. Previewers are offered bets on some future performance measures of the movie, like the Rotten Tomatoes rating or the number of tickets sold, versus those of another movie of the same type. To put it simply, our mechanisms ask people to bet on the *relative* performance of the previewed movie. Doing so alleviates the concern of Keynesian beauty-contest type of herding, when agents act upon what they think others will think, rather than upon their own signal. With a betting mechanism on absolute performance, as in a prediction market, agents' decisions are jointly determined by their private signals and their prior expectations about movie performance. Betting on relative performance, as in our mechanisms, disentangles the private signal from prior expectations, as we will show.

This paper introduces simple betting mechanisms (Top-Flop betting and Threshold betting) and determines sufficient conditions for the chosen bets to reveal private signals. The first part of the paper considers a setting where a single agent receives a signal about one item and bets on its rating relative to that of another item belonging to a collection of similar items. In this setting, we assume that the ratings are exogenous random variables. There are two key conditions for the agent to reveal his signal through his betting behavior. First, the rating of an item must be more informative about the signals related to that item than the ratings of other items are. For instance, learning that the previewed movie grossed more than $500M on its first weekend is more informative about the probability to like that specific movie than learning that another movie exceeded the same milestone is. Second, the agent has the same prior for all items of the collection. That is, the agent has no reason to prefer one movie over the other ex ante. Our results do not require the agent to be risk neutral (or even a risk-averse expected-utility maximizer), but simply to choose the bet giving the higher chance to win. Hence, our results are valid for any decision model satisfying first-order stochastic dominance.

In the second part of the paper, we consider a game setting with at least four agents and provide a theoretical foundation for the rating. For a given agent, the rating for an item in the collection is determined by betting choices of other agents. Similarly to the single-agent case, each agent in a betting game receives a signal about one item in the collection. We again establish sufficient conditions for agents to reveal their signals. Specifically, we do not require that they fully agree on how signals are generated and how signals of any two agents are related. Agents may think they all have a different prior probability to like a given movie. They may even disagree about what these probabilities are. They do agree that the signals of two agents are more positively correlated when the signals are for the same item than for different items. However, they may disagree on the exact degree of correlation. The results we obtain are partial implementation results. We establish that agents revealing their signal is a Nash equilibrium but other equilibria are not excluded.

Several methods have been proposed to reveal unverifiable signals in survey settings (Prelec, 2004; Witkowski and Parkes, 2012*b*; Radanovic and Faltings, 2013; Baillon, 2017; Cvitanić et al., 2019). They provide truth-telling incentives by asking each agent two questions regarding a single item. One of the questions is directly about the signal and the other one is about predicting other agents' answers. These methods are based on a common-prior assumption, requiring that agents only differ in the signal they received. With these methods, truthful signal reporting is a Bayesian Nash equilibrium when agents are risk neutral. By using more than one item, we can relax the common prior assumption and replace it by an assumption about how the items are related. In other words, in our model, priors may differ across agents but have to agree across items.

Witkowski and Parkes (2012*a*) also introduced a method that relaxed the common prior assumption but it required to elicit priors before agents receive their signals. We do not require such additional elicitation. In that sense, our mechanism is *minimal*, as defined by Witkowski and Parkes (2013). The latter paper proposes a minimal mechanism approximating beliefs with the empirical distribution of signals and delaying payment until the distribution is accurate enough. We do not need such delays. Our approach also allows us to use a payment rule that is simpler than the aforementioned mechanisms and is robust to risk aversion, certainty effects, and other behavioral phenomena. Finally, the game-theoretic version of our mechanisms is based on assumptions that are close to those of Dasgupta and Ghosh (2013) and Shnayder et al. (2016). These authors also used cross-item correlations to incentivize truthful signal reporting (including non binary signals for Shnayder et al., 2016), but they needed that all agents get signals for at least two items. The literature is further discussed in Section 5.4.

We conclude our paper with examples of practical implementations and potential applications of our methods. We show how Threshold betting can be implemented as a financial derivative (an option) of prediction markets. We also explain how our simple bets can be used to assess whether people are willing to pay a given amount for product features that are yet to be developed.

5.2 Betting on exogenous ratings

5.2.1 Signals, ratings, and beliefs

We first consider a setting of a single agent ("he"). There is a *collection of items* $\mathcal{K} \equiv \{1, \dots, K\}$ with $K \geq 2$. For one[2] fixed $l \in \mathcal{K}$, the agent receives a private signal, modeled as a realization $t \in \mathcal{T} = \{0,1\}$ of a random variable T. A *center* ("she") wishes to elicit t. For instance, $\mathcal{K}$ is a collection of movies, the agent watches movie l, and the center wants to know whether he liked it ($t = 1$) or not ($t = 0$). Each item $k \in \mathcal{K}$ has a rating, reflecting its quality and taking values from $\mathcal{S}$, a countable subset of the reals. The ratings are unknown to the agent and to the center when the agent receives t. Furthermore, neither the agent nor the center can influence the ratings. Hence, ratings are modeled as bounded[3] random variables Y_k with generic realization $y_k \in \mathcal{S}$.

We assume that all the random variables (ratings and signals) are defined on the same probability space $(\Omega, \mathcal{F}, P)$. By Kolmogoroff (1933), this can always been assumed. For simplicity, we avoid measure-theoretic complications and assume that Ω is countable, that $\mathcal{F}$ is the sigma-algebra of all subsets of Ω (called *events*), and that P is countably additive.[4] The random variables (and P) need not describe some objective processes but rather the agent's beliefs. His prior probability to get signal 1 is $P(t = 1)$ and H_k denotes the distribution function of his prior about the rating.

Assumption 5.1 (Identical prior). *For any $k \in \mathcal{K} \setminus \{l\}$, Y_k and Y_l are identically distributed, with $H_k = H_l$.*

[2] We assume that, if the agent receives signals about other items, the corresponding items are removed from the collection and that the assumptions introduced below hold conditional on the additional signals.

[3] A real-valued random variable $Y_k = Y_k(\omega)$ defined on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is bounded if there exists a constant M such that $| Y_k(\omega) | \leq M$ for all $\omega \in \Omega$.

[4] For instance, Ω may be the Cartesian product of the rating space and the signal space, $\Omega = (\Pi_{k \in \mathcal{K}} \mathcal{S}) \times \mathcal{T}$.

Let H ($\equiv H_l$) be the prior, identical for all items, as defined in Assumption 5.1. Assumption 5.1 means that the agent has the same expectations about the items in the collection before he receives a signal about item l. In practice, it requires that items are similar. In the movie example, if the rating is a performance measure such as reviews or gross revenue, the collection should not mix blockbusters with independent movies because the agent may have very different expectations of the rating for the two categories. Dasgupta and Ghosh (2013) and Shnayder et al. (2016) argued for the identical prior assumption when the agent is ignorant about the collection and items are randomly assigned. They typically considered agents completing multiple tasks that are crowd-sourced, such as image labeling, peer-assessment in online courses, or reporting features of hotels and restaurants.

A subset of the rating space, useful for what follows, is $\mathcal{S}' = \{y \in \mathcal{S} : 0 < H(y) < 1\}$. It excludes all ratings that are so low or so high that the agent believes they will never occur. It also excludes the maximum rating level the agent believes may occur (the smallest y such that $H(y) = 1$).[5] We consider the non-trivial case where the agent believes that more than one rating level may occur, i.e. $\mathcal{S}'$ not empty.

Assumption 5.2 (Comparative informativeness)**.** *For all $k \in \mathcal{K} \setminus \{l\}$ and $y \in \mathcal{S}$, $P(t = 1 \mid Y_l > y) > P(t = 1 \mid Y_k > y)$.*

In the mechanism design literature, private signals are linked to states of nature by a signal technology. Here, the possible ratings play the role of the states of nature. The signal technology is (believed by the agent to be) such that that the rating of item l is more positively associated with receiving a signal 1 about l than the rating of item k is.[6] Let the collection of items be, for instance, all movies of a franchise, and the rating be how much the movies will earn in the first month after their release. If the agent learns that movie $l = 4$ has grossed \$20,000,000 so far (so Y_4 will be at least that amount), he may update his probability of liking that movie upwards. If instead, he learns that another movie, e.g. $k = 3$, has grossed \$20,000,000 so far, he may also update his probability to like movie 4 upwards but less so. He may even decrease his probability to like movie 4 if he thinks that a great movie 3 means a less good movie 4. Our assumption allows for biases or distrust of the underlying ratings. For instance, the

[5] $\mathcal{S}'$ does not coincide with the support of the distribution. For instance, if $\mathcal{S} = \{1, \ldots, 6\}$ and the support is $\{2,4,5\}$, then $\mathcal{S}' = \{2,3,4\}$. It excludes the highest value of the support, 5, but includes 3 because $0 < H(3) < 1$ even though $P(Y_k = 3) = 0$. We use $\mathcal{S}'$ because, as will become transparent later, our mechanisms rely on properties of cumulative distribution functions, not probability (or density) functions.

[6] Assumption 5.2 also implies $P(t = 1) \in (0, 1)$ because a degenerate prior would give the same posterior no matter what Y_l and Y_k would be.

agent may think that the rating is biased by the fact that some people see all movies of the franchise anyhow, good or bad, as long as the biases neither eliminate nor reverse the stronger relation between a high rating of l and a signal 1 than between a high rating of k and a signal 1.

Once the agent learns his signal t, he updates his beliefs about the ratings, which yields the posterior distribution function $F_k^t(y) = P(Y_k \leq y \mid T = t)$. Assumptions 5.1 and 5.2 guarantees that the signal influences his expectations about Y_l in a very specific way relatively to any other Y_k. For any two cumulative distribution functions F and G with domain $\mathcal{S}$, we write $F \succeq_{SD} G$ ($F \succ_{SD} G$) and say that F *(strictly) first-order stochastically dominates* G when $F(y) \leq G(y)$ for all $y \in \mathcal{S}$ (with $F(y) < G(y)$ for some y).

Lemma 5.1. *Assumptions 5.1 and 5.2 imply* $F_l^1(y) \succ_{SD} F_k^1(y)$ *and* $F_k^0(y) \succ_{SD} F_l^0(y)$ *for all* $k \neq l$*.*

The proof of Lemma 5.1, as all other proofs, is in Appendix. Intuitively, a signal $t = 1$ is more associated with high ratings of item l than with high ratings of item k and therefore shifts more posterior F_l^1 to the right than posterior F_k^1. Note that we could have immediately assumed the implications of Lemma 5.1, which would be more general than Assumptions 5.1 and 5.2. The advantage of providing sufficient conditions is to clarify what types of items and ratings can be used. If the agent believes the rating of l is more positively correlated with the signal than the rating of k is and views all items of the collection as equivalent, ex ante, in terms of ratings, then his beliefs about the ratings of l and of any $k \neq l$ once he has received his signal will satisfy the stochastic dominance properties spelled out in Lemma 5.1. These properties guarantee that signals can be identified from beliefs. Before we design bets based on this identification strategy, we introduce an additional assumption that we will use in some of our results, in which we need the random variables Y_k and Y_l to be not only identically distributed but also independent.

Assumption 5.3 (Independence). *For any* $k \in \mathcal{K}$ *with* $k \neq l$*,* Y_k *and* Y_l *are independent, and conditionally independent given* T*.*

We could also replace conditional independence in Assumption 5.3, using the fact that Y_k and Y_l are independent, by:

$$\frac{P(t = 1 \mid Y_l, Y_k)}{P(t = 1 \mid Y_l)} = \frac{P(t = 1 \mid Y_k)}{P(t = 1)}. \tag{5.1}$$

In other words, how information about Y_k changes the probability of a positive signal is invariant to information about Y_l.

5.2.2 The bets

Let π be a *prize* (money, a gift, or... an actual pie) that the agent likes. The absence of prize is denoted by 0. Let $\mathcal{E}$ be an event, an element of $\mathcal{F}$. A *bet on* $\mathcal{E}$ assigns π to $\mathcal{E}$ and 0 to the complement of $\mathcal{E}$. The agent has preferences over bets. If we do not explicitly mention that preferences are strict, we mean weak preferences.

Assumption 5.4 (Probabilistic sophistication). *For any three events $\mathcal{E}$, $\mathcal{E}'$, and $\mathcal{G} \in \mathcal{F}$, the agent prefers a bet on $\mathcal{E}$ to a bet on $\mathcal{E}'$ when he knows that $\mathcal{G}$ occurred if and only if $P(\mathcal{E} \mid \mathcal{G}) \geq P(\mathcal{E}' \mid \mathcal{G})$.*

Assumption 5.4 says that the agent is probabilistically sophisticated in the sense of Machina and Schmeidler (1992), and furthermore, that preferences are consistent with P, the (subjective) probability measure that underlies the random variables. He may be risk neutral, or be a risk-averse expected utility maximizer, or even transform his probabilities as long as the transformation is strictly increasing in P so as to satisfy stochastic dominance (Kahneman and Tversky, 1979; Tversky and Kahneman, 1992). Assumption 5.4 implies that the agent strictly prefers π (a bet on Ω) to nothing (a bet on $\varnothing$).

Definition 5.1. *For an arbitrary $k \in \mathcal{K} \setminus \{l\}$, a* Top bet *is a bet on $\{\omega \in \Omega : Y_l(\omega) > Y_k(\omega)\}$ and a* Flop bet *is a bet on $\{\omega \in \Omega : Y_l(\omega) < Y_k(\omega)\}$.*

The center proposes a Top bet and a Flop bet to the agent, who may choose one of them (or reject both).

Lemma 5.2. *Under Assumptions 5.1 to 5.4, the agent, before learning t, is indifferent between the Top and the Flop bet but strictly prefers any of them to nothing.*

Ex ante, the agent has the same belief H about the distribution of Y_k and Y_l (Assumption 5.1), which are also independent (Assumption 5.3), and there is no reason to prefer betting on one rating being higher rather than the other (Assumption 5.4). Furthermore, the agent does not expect the ratings to be equal with certainty, and therefore expects that both bets have a nonnull chance to yield the prize. The agent wants to participate in the betting. When he learns his signal, he has a clear preference for one of the bets, as established by the next Theorem.

Theorem 5.1. *Under Assumptions 5.1 to 5.4, for any $k \in \mathcal{K} \setminus \{l\}$, the agent strictly prefers the Top bet if $t = 1$ and the Flop bet if $t = 0$.*

The following corollary makes explicit that the agent does not need to know k, which can be selected from the collection of items with a random device. We assume, here and whenever we will refer to such exogenous random devices, that they are independent of all the random variables described so far and also conditionally independent given T, and that all elements of the collection have a positive probability to be drawn.

Corollary 5.1. *Theorem 5.1 remains valid if k is unknown to the agent and, instead, will be randomly drawn from $\mathcal{K} \setminus \{l\}$.*

Even though the agent does not know which k will be drawn from item collection $\mathcal{K}$, the collection should still be clearly specified. If the agent can imagine any item, Assumptions 5.1 to 5.3 are less likely to hold.

Our results for the Top and Flop bets rely on (conditional) independence of the ratings. The center can also propose another type of simple bets to the agents, which still reveal signals but without relying on independence, only on the stochastic dominance conditions established in Lemma 5.1. For instance, the agent could be asked to bet on whether the rating of item l or the rating of item k will exceed some threshold. We call this approach *Threshold betting*.

Definition 5.2. *A* Threshold-y bet *on k is a bet on $\{\omega \in \Omega : Y_k(\omega) > y\}$.*

If the ratings are taken from Rotten Tomatoes, a Threshold-60 bet would yield the prize only if the rating of the movie exceeds 60%. Ex ante, the agent is indifferent between the items on which the Threshold-y bets are based.

Lemma 5.3. *Under Assumptions 5.1 and 5.4, for any $y \in \mathcal{S}'$ and $k \in \mathcal{K} \setminus \{l\}$, the agent, before learning t, is indifferent between a Threshold-y bet on k and a Threshold-y bet on l, but strictly prefers any of them to nothing.*

Assumptions 5.1 to 5.4 are about the agent's beliefs and behavior, and not about objective features of a signal technology. In that sense, they may be difficult to verify. However, Lemma 5.3 provides a way to jointly test 5.1 and 5.4. Before previewing a movie, the agent should be indifferent between the bets.

Theorem 5.2. *Under Assumptions 5.1, 5.2, and 5.4, for any $y \in \mathcal{S}$ and $k \in \mathcal{K} \setminus \{l\}$, the agent strictly prefers a Threshold-y bet on l to a Threshold-y bet on k if $t = 1$ and a Threshold-y bet on k to a a Threshold-y bet on l if $t = 0$.*

Corollary 5.2. *Theorem 5.2 remains valid if k is unknown to the agent and will be randomly drawn from $\mathcal{K} \setminus \{l\}$ and/or if y is unknown to the agent and will be randomly drawn from $\mathcal{S}$.*

A challenge of Theorem 5.2 is to find a value from the support to use as threshold, because the support, unlike the domain, is subjective. The center can mitigate the problem by avoiding extreme values. Corollary 5.2 solves the challenge by proposing to randomly draw a value from $\mathcal{S}$ after the agent chooses a bet.

Before receiving a signal, the agent is indifferent between Top and Flop bets (Lemma 5.2) and also between Threshold-y bets on l and Threshold-y bets on k (Lemma 5.3). No matter which signal he receives, his winning probability always increases if he chooses optimally. With Threshold-y bets, the winning probability with optimal choices is $P(t=1)P(Y_l > y \mid t=1) + P(t=0)P(Y_k > y \mid t=0)$, which strictly exceeds the no-signal chance of winning $P(Y_l > y)$ $(= P(Y_k > y))$.[7] The difference between the two gives us the ex ante value of the signal (in terms of winning chances). The same reasoning applies to Top-Flop betting.

Now imagine that the agent has to pay a cost (or provide an effort) to acquire the signal. He will compare this cost to the benefit–the increase in the probability of getting π.

Remark 5.1. *The ex ante value of the signal is positive. Hence, under common regularity assumptions (continuity in utility), there exists a non-degenerate range of costs that the agent is willing to pay to acquire the signal.*

How much (effort) the agent is willing to spend on the signal will depend on his whole utility function. Calculating it would require specifying further assumptions about the decision model of the agent (beyond Assumption 5.4). Obviously, we can expect that increasing the value of the prize will increase the maximum cost the agent is willing to pay. What we claim is that our simple bets can stimulate signal acquisition. In practice, they can be used to motivate people to look for a piece of information, preview a movie, or carefully evaluate a product.[8]

[7] Proof: $P(Y_l > y) = P(t=1)P(Y_l > y \mid t=1) + P(t=0)P(Y_l > y \mid t=0)$ by definition. Replacing the $P(Y_l > y \mid t=0)$ by the strictly larger $P(Y_k > y \mid t=0)$ (according to Theorem 5.2) establishes the result.

[8] If the incentives are too high, the approach can backfire and the agent may start looking for other pieces of information than his private signal, distorting what the center aimed to elicit. In the context of belief elicitation with scoring rules, this problem has been discussed by Schotter and Trevino (2014) and a solution has been proposed by Tsakas (2020).

5.3 Betting on endogenous ratings

5.3.1 Agents, their signals, and their beliefs

We now consider multiple agents $i \in \mathcal{I} = \{1, \cdots, Kn\}$, i.e., $n \geq 2$ agents per item. In the simplest case, with two items, we need a minimum of 4 agents. In this section, most variables and objects from the previous section become agent-specific, which will be indicated by subscript i. Each agent i gets a signal $T_i \in \mathcal{T} = \{0, 1\}$, about item $l_i \in \mathcal{K}$. The set of agents with a signal about k is $\mathcal{I}_k \equiv \{j \in \mathcal{I} : l_j = k\}$ and it has cardinality n. The state space is $\Omega = \mathcal{T}^{Kn}$, where a state ω is the vector of signals received by the Kn agents. (We need not specify ratings here, as will become apparent later.)

Agent i will be offered to bet on ratings based on the others' actions in the games to be defined in the next subsection. For item $k = l_i$, "the others" mean $\mathcal{I}_{i,k} \equiv \mathcal{I}_k \setminus \{i\}$. In what follows, it will be desirable to consider sets of agents with the same cardinality as this set of others. We, therefore, define for items $k \neq l_i$, $\mathcal{I}_{i,k} \equiv \mathcal{I}_k \setminus \{j\}$ with $j = \max \mathcal{I}_k$ (any other j could have been chosen as well). We can now define the analog of the random variables Y_k of the preceding section. For all i and k,

$$Y_{i,k} = \sum_{j \in \mathcal{I}_{i,k}} T_j. \tag{5.2}$$

The random variable $Y_{i,k}$ is, for agent i, the number of other agents who received signal 1 for item k. As in the previous section, agent i's belief P_i, defined over Ω, generates distribution priors $H_{i,k}$ about $Y_{i,k}$. The domain of $H_{i,k}$ is $\mathcal{S}_i = \mathcal{S} = \{0, \ldots, n-1\}$ because $Y_{i,k}$ can take values between 0 and $n-1$. The sets $\mathcal{S}_i'$ is defined similarly as $\mathcal{S}'$ in the preceding section.

Example 5.1. *The simplest case of our setting is $n = K = 2$, involving four agents. State ω is a quadruplet of signals (t_1, t_2, t_3, t_4). With $l_1 = l_2 = 1$, $l_3 = l_4 = 2$, $\mathcal{I}_{1,2} = \{3\}$, and $\omega = (t_1, t_2, t_3, t_4)$, we have $Y_{1,1}(\omega) = t_2$ and $Y_{1,2}(\omega) = t_3$.*

Assumption 5.5 (Common knowledge)**.** *Agents share the common belief that Assumption 5.4 holds for all agents $i \in \mathcal{I}$, with all P_is themselves common knowledge.*

Assumption 5.5 means that agents may all have different P_i but they know that everyone satisfies first order stochastic dominance with respect to their own beliefs. Furthermore, if Assumptions 5.1, 5.2, and 5.3 hold for all P_i, then this fact is automatically common knowledge because the beliefs P_is are themselves common knowledge. Assumptions 5.1, 5.2, and 5.5 do not require that all agents in $\mathcal{I}_k$ have the same probability to get a signal 1. Agent i can think everyone is different, and even that some people dislike everything (trolls). What we need is that each agent i perceives T_i and $Y_{i,k}$ more associated when $k = l_i$ than when $k \neq l_i$. Independence (Assumption 5.3) can now be justified if, for instance, signals of any two agents i and j are independent when $l_i \neq l_j$.

5.3.2 The games

In what follows, we will first define interim preferences, i.e. preferences conditional on signals: what agents believe and prefer if their signal is 0 versus if their signal is 1. Agents must then decide, ex ante, what they will do for each possible signal. We will obtain a Bayesian game and, finally, define a (Bayesian) Nash equilibrium of this game.

We first define a generic game with the same *action set* $\mathcal{A} = \{0,1\}$ for all agents, with a_i the action of agent i. The *payoff function* of the game for agent i is $\Pi_i : \mathcal{A}^{Kn} \longrightarrow \{0,\pi\}$. Each agent chooses a *strategy*, which is a pair of actions $(a_i^0, a_i^1) \in \mathcal{A}^2$, where a^0 will be implemented in state ω if $T_i(\omega) = 0$ and a^1 will be implemented if $T_i(\omega) = 1$. A *strategy profile*, i.e. the strategy of all agents, is denoted by $(a^0, a^1) \in \left(\mathcal{A}^2\right)^{Kn}$. The *implemented action* for agent i in state ω is $a_i^{T_i(\omega)}$, which we write a_i^ω for short. We similarly denote $a^\omega \in \mathcal{A}^{Kn}$ the *profile of implemented actions*.

Example 5.1 (continued). *A strategy profile is of the form* $((a_1^0,a_1^1),(a_2^0,a_2^1),(a_3^0,a_3^1),(a_4^0,a_4^1))$. *If the realized state is* $\omega = (0,1,1,0)$, *then the profile of implemented actions is* $a^\omega = (a_1^0, a_2^1, a_3^1, a_4^0)$. *The payoff function* Π_i *of agent i assigns either 0 or* π *to any such quadruplet.*

The agents have (interim) preferences over strategy profiles, conditional on their signal and denoted by $\succsim_{i|T_i}$. Assumption 5.5, which includes Assumption 5.4, implies that it is common knowledge that $(a^0,a^1) \succsim_{i|T_i} (b^0,b^1)$ if and only if

$$P_i\left(\{\omega \in \Omega : \Pi_i\left(a^\omega\right) = \pi\} \mid T_i\right) \geq P_i\left(\{\omega \in \Omega : \Pi_i\left(b^\omega\right) = \pi\} \mid T_i\right). \tag{5.3}$$

In Equation 5.3, the agent first determines which are the states ω yielding π if the strategy profile is (a^0,a^1), and if the strategy profile is (b^0,b^1). The agent then compares the probability (given his signal) of the states yielding π when the strategy profile is (a^0,a^1) to the probability obtained if the strategy profile is (b^0,b^1). Agent i finally chooses the strategy profile that gives the higher chance to get π.

With $\mathcal{I}$, Ω, $\mathcal{A}$, $\mathcal{T}$, T_i, P_i, and $\succsim_{i|T_i}$, we have defined a Bayesian game, further assuming common knowledge of Ω, $\mathcal{I}$, $\mathcal{T}$, $\mathcal{A}$, and the Π_is.[9] Let $(b_i^0, b_i^1; a^0, a^1)$ be the strategy profile, which replaces a_i^0 and a_i^1 by b_i^0 and b_i^1 in (a^0,a^1). A strategy profile (a^0,a^1) is

[9] Harsanyi (1968) defined Bayesian games where the difference in beliefs arise from an objective information mechanism, which is common knowledge. Interim beliefs may differ but prior beliefs are the same. In our case, prior beliefs may also differ. However, the (possibly different) priors are common

a *Nash equilibrium* of the Bayesian game if for all $i \in \mathcal{I}$, $(a^0,a^1) \succsim_{i|T_i} (b_i^0,b_i^1;a^0,a^1)$ for all $(b_i^0,b_i^1) \in \mathcal{A}^2$. We say that the Nash equilibrium is *strict* if, in addition and for all i, $(a^0,a^1) \succ_{i|T_i=0} (b_i^0,a_i^1;a^0,a^1)$ for all $b_i^0 \in \mathcal{A} \setminus \{a^0\}$ and $(a^0,a^1) \succ_{i|T_i=1} (a_i^0,b_i^1;a^0,a^1)$ for all $b_i^1 \in \mathcal{A} \setminus \{a^1\}$. Strict means that the implemented action is strictly preferred (even though the not-implemented action is only weakly preferred).

We can now define Top-Flop and Threshold-y games. Each agent i will be offered bets on (individualized) ratings $\widehat{Y}_{i,k}$ defined as a function of an action profile $a \in \mathcal{A}^{Kn}$ by:

$$\widehat{Y}_{i,k} = \sum_{j \in \mathcal{I}_{i,k}} a_j. \tag{5.4}$$

In Section 5.2, the ratings were exogenous and agents had beliefs about them. In the present section, we provide a game-theoretic foundation for the ratings, which are endogenously defined by the actions of others. Agents now have beliefs about signals, which translate into beliefs about ratings $\widehat{Y}_{i,k}$ for a given strategy profile. The payoff function of the game is defined on the $\widehat{Y}_{i,k}$s. We first assign h_i to each agent i, given by $h_i = l_i + 1$ if $l_i < K$ and $h_K = 1$.

Definition 5.3. *In a* Top-Flop game, *Π_i assigns π to $\left\{a \in \mathcal{A}^{Kn} : a_i = 1 \;\&\; \left(\widehat{Y}_{i,l_i} > \widehat{Y}_{i,h_i}\right)\right\}$ (Top case) and to $\left\{a \in \mathcal{A}^{Kn} : a_i = 0 \;\&\; \left(\widehat{Y}_{i,l_i} < \widehat{Y}_{i,h_i}\right)\right\}$ (Flop case). It assigns 0 to all other elements of $\mathcal{A}^{Kn}$.*

The payoff function is defined such that choosing action 1 is equivalent to choosing a Top bet; it pays π if $\widehat{Y}_{i,l_i} > \widehat{Y}_{i,h_i}$. Similarly, choosing action 0 is equivalent to choosing a Flop bet, which pays off if $\widehat{Y}_{i,l_i} < \widehat{Y}_{i,h_i}$.

Example 5.1 (continued). *With $l_1 = l_2 = 1$, $l_3 = l_4 = 2$, agents 1 and 2 get a signal about item 1, and agents 3 and 4 get a signal about item 2. Furthermore, $\widehat{Y}_{1,1} = a_2$ and $\widehat{Y}_{1,2} = a_3$, which means agent 1 bets on the actions of agents 2 and 3. The following table describes Π_1.*

$\widehat{Y}_{1,1}$	$\widehat{Y}_{1,2}$	$a_1 = 0$	$a_1 = 1$
$a_2 = 0$	$a_3 = 0$	*0*	*0*
$a_2 = 0$	$a_3 = 1$	π	*0*
$a_2 = 1$	$a_3 = 0$	*0*	π
$a_2 = 1$	$a_3 = 1$	*0*	*0*

knowledge, which still allows agents to infer others' interim beliefs and preferences. See Osborne and Rubinstein (1994, Section 2.6.3.) for a discussion, and their Definition 25.1 of a Bayesian game and Definition 26.1 of a Nash equilibrium of a Bayesian game, which we followed here.

First note that for agent 1, the action of agent 4 does not affect his payment. Second, he wins π in two cases: (i) if he and agent 2 report 0 while agent 3 reports 1 and (ii) if he and agent 2 report 1 while agent 3 reports 0. Case (i) is a Flop bet, where item 2 gets a higher rating ($\widehat{Y}_{1,2} = 1$) than item 1 ($\widehat{Y}_{1,1} = 0$). Symmetrically, case (ii) is a Top bet.

Theorem 5.3. *If all agents $i \in \mathcal{I}$ satisfy Assumptions 5.1 to 5.4, and if Assumption 5.5 holds, then (a^0, a^1) with $a_i^0 = 0$ and $a_i^1 = 1$ for all $i \in \mathcal{I}$ is a strict Nash equilibrium of a Top-Flop game.*

In the proof (Appendix 5.B), we first establish that if every $j \neq i$ plays $(0,1)$, then $\widehat{Y}_{i,k} = Y_{i,k}$ for all k. By Theorem 5.1, the best response of agent i is then to choose a Flop bet if $T_i = 0$ and a Top bet if $T_i = 1$, hence picking strategy profile $(0,1)$. All this is common knowledge, so the agents' beliefs are consistent with the Nash equilibrium.

Corollary 5.3. *Under the assumptions of Theorem 5.3, all agents strictly prefer the equilibrium of a Top-Flop game in which all agents play $(0,1)$ to all agents playing $(0,0)$ or all agents playing $(1,1)$.*

By construction, degenerate strategy profiles where everyone plays $(0,0)$ or everyone plays $(1,1)$ yields payoff 0. Hence, the equilibrium $(0,1)$ is preferred because it gives a chance to get π. We now turn to Threshold-y betting that we similarly transform into a game.

Definition 5.4. *In a* Threshold-y game, *for $y \in \{0, \dots, n-2\}$, Π_i assigns π to*
$\left\{a \in \mathcal{A}^{Kn} : a_i = 1 \;\&\; \left(\widehat{Y}_{i,l_i} > y\right)\right\}$ *and to*
$\left\{a \in \mathcal{A}^{Kn} : a_i = 0 \;\&\; \left(\widehat{Y}_{i,h_i} > y\right)\right\}$ *. It assigns 0 to all other elements of $\mathcal{A}^{Kn}$.*

With the payoff functions of a Threshold-y game, agent i gets π when playing 1 if item l_i exceeds threshold y and when playing 0 if item h_i exceeds threshold y. The threshold can be any value up to $n-1$ because $\widehat{Y}_{i,k}$ can never exceed n.

Example 5.2 (continued). *With four agents, only a Threshold-0 game is possible.*[10] *Agent 1 still bets on the actions of agents 2 and 3 but Π_1 is now:*

$\widehat{Y}_{1,1}$	$\widehat{Y}_{1,2}$	$a_1 = 0$	$a_1 = 1$
$a_2 = 0$	$a_3 = 0$	*0*	*0*
$a_2 = 0$	$a_3 = 1$	π	*0*
$a_2 = 1$	$a_3 = 0$	*0*	π
$a_2 = 1$	$a_3 = 1$	π	π

[10] $\widehat{Y}_{i,k}$ can only be 0 or 1, and therefore can only strictly exceed 0.

Agent 1 wins π in two cases: (i) if he and agent 2 play 1 ($a_1 = a_2 = 1$) and (ii) if he plays 0 while agent 3 plays 1 ($a_1 = 0$ and $a_3 = 1$). Case (i) is a bet on the rating of item 1 (= the action of agent 2) exceeding 0 and case (ii) a bet on the rating of item 2 (= the action of agent 3) exceeding 0. The last row of the table differs from the Top-Flop game.

Theorem 5.4. *If all agents $i \in \mathcal{I}$ satisfy Assumptions 5.1, 5.2, and 5.4, and if Assumption 5.5 holds, then (a^0, a^1) with $a_i^0 = 0$ and $a_i^1 = 1$ for all i is a strict Nash equilibrium of a Threshold-y game when $y \in \mathcal{S}_i^l$ for all i.*

Corollary 5.4. *Under the assumptions of Theorem 5.4, (a^0, a^1) with $a_i^0 = 0$ and $a_i^1 = 1$ for all i is a strict Nash equilibrium of a Threshold-y game when y is randomly drawn from $\mathcal{S}$.*

Theorem 5.4 has two main limitations. First, all agents must think the threshold is not trivial, neither too high nor too low. A solution, given by Corollary 5.4 is to randomly draw the threshold ex post. Second, unlike in the Top-Flop game, there exists an equilibrium that would be preferred by all agents to playing $(1,0)$. If they all play $(1,1)$, they can all win with certainty. This equilibrium can be excluded by altering Π_i such that it is 0 if $\widehat{Y}_{i,l_i} = \widehat{Y}_{i,h_i} = n-1$ (the maximum rating). This modification of the payoff function is not anodyne though, and requires to bring back Assumption 5.3.[11]

5.4 Discussion

5.4.1 Limitations and related literature

In the exogenous-rating setting, it is important that the agent does not expect the center to have control over Y_k. A suspicious agent would then enter a game with the center. Suspicion can be avoided or at least mitigated by using ratings controlled by an independent third party or involving a multitude of people. For instance, the rating can be the price established on a large prediction market at a given time. This would make clear that influencing the rating would cost more to the center than paying π to the agent.

Our exogenous-rating setting relates to the literature on canonical contract design for adverse selection problems as in Mirrlees (1971), Maskin and Riley (1984) and Baron and Myerson (1982). For instance, in the classical monopoly setting, the principal (the center in our setting) does not know the agent's private information, but she can screen different types of agents by offering them an incentive compatible menu of contracts,

[11] The probability of getting π does not depend anymore on either $\widehat{Y}_{i,l_i}$ if $a_i = 1$ or $\widehat{Y}_{i,h_i}$ if $a_i = 0$, but on both $\widehat{Y}_{i,l_i}$ and $\widehat{Y}_{i,h_i}$ for all a_i.

under which the agent will pick the one revealing his true type. Since the screening is achieved by leveraging the structure of agents' preferences, the principal is required to know the preference for each type and its distribution. Our methods do not require that because our screening techniques are mainly based on the complementarity between the rating and the private signal for each agent. This is possible because, in our setting, agents have no other incentives (to either reveal or hide the signals) than trying to win the prize.

Our Bayesian game setting relates to a strand of literature in mechanism design, including Myerson (1986) and Crémer and McLean (1988). Both mechanisms construct truth-telling equilibrium by exploiting the correlation of private information across agents. As in Myerson (1986), truth-telling in our paper is an equilibrium, but need not be the only one. Hence, undesirable equilibria may also occur and our Theorems 5.3 and 5.4 are partial implementation results. By contrast, Maskin (1999) constructed mechanisms with full implementation, i.e., not only admitting desirable equilibria but also excluding undesirable ones. Unlike in Crémer and McLean (1988), the person extracting the information (the center) in our setting does not need to know the prior of the agents. Our mechanisms are detail-free; they can be implemented without knowing the details of the signal technology. In that sense, the Top-Flop and Threshold games get very close to the desiderata of the Wilson's doctrine (Wilson, 1987).

More recently, Bergemann and Morris spurred a renewed interest in partial and full implementation problems that do not rely on strong assumptions about agents' beliefs (Bergemann and Morris, 2005, 2009*b*,*a*). This led to the literature on robust implementation. Our results do not attain robustness in the sense that they do not guarantee incentive compatibility for all possible beliefs. They allow, however, for a relatively rich set of beliefs under common knowledge Assumption 5.5. Our approach in that regards is closest to that of Ollár and Penta (2017) and Ollár, Penta et al. (2019), who studied partial and full implementation under sets of beliefs based on common knowledge assumptions. Assumption 5.5 is an instance of the general belief restrictions in Ollár and Penta (2017).

Bayesian methods to elicit private signals in surveys or on crowd-sourcing platforms have been proposed by Prelec (2004), Miller, Resnick and Zeckhauser (2005), Witkowski and Parkes (2012*b*), Radanovic and Faltings (2013), Baillon (2017), and Cvitanić et al. (2019). All these papers rely on common prior assumptions, sometimes weakly relaxing them. Our common knowledge assumption is much weaker, allowing all agents to disagree on the probability to observe some signals. Note that for the Nash equilibrium to be credible, the key point is not so much that agents know the priors of all other agents but rather that they know that these priors are well-behaved as described by Assumptions 5.2 to 5.3.

Witkowski and Parkes (2013) were first to show that using multiple tasks relaxes the common prior and allows for beliefs to diverge from some 'true' signal technology. They provide a mechanism that is minimal, like ours, and unlike the papers discussed in the previous paragraph with the exception of Miller, Resnick and Zeckhauser (2005), in that it only requires one report (in our case, one bet) from each agent. Their mechanism then uses the empirical signal distribution, to be elicited over time, as a proxy for beliefs and applies a scoring approach comparable to that of Miller, Resnick and Zeckhauser (2005). Our mechanisms do not require such payment delays, and our payoff rules are simpler and more transparent than theirs.

Our beliefs assumptions are very close to those of Dasgupta and Ghosh (2013) and Shnayder et al. (2016). These papers consider a signal correlation matrix and assume that it describes the beliefs of all agents. However, Shnayder et al. (2016) do point out that only the structure of the correlations matters and therefore heterogeneity in beliefs would be possible (their footnote 7 and subsection 5.4). Unlike the present paper, Dasgupta and Ghosh (2013) and Shnayder et al. (2016) only consider game settings and require that each agent receives signals about two items (in their setting, performs two tasks) whereas our agents receive a signal about only one item.

A major limitation of our paper, which is shared by Dasgupta and Ghosh (2013) but not by Shnayder et al. (2016), is that we can only handle binary signals. Extending our results to non-binary signals is not trivial and would require much heavier assumptions about beliefs, especially correlations between signals and ratings. With binary signals, signal 1 being associated with high ratings means that signal 0 is associated with low ratings. With non-binary signals, such implications do not hold anymore. Imagine that signals are satisfaction levels $\{1,2,3\}$ and that we have, for each item k, three ratings Y_k^1, Y_k^2, and Y_k^3 (for instance, the number of other agents reporting signals 1, 2, and 3 respectively). An agent with satisfaction level 3 can reasonably increase the probability that Y_k^3 is at least y but also the probability that Y_k^2 is at least y. A possible approach is to split the agent sample between three groups. Some agents get the possibility to bet on Y_k^3 vs Y_l^3, which can reveal whether their signal was 3 or not 3. Other agents get the possibility to bet on Y_k^2 vs Y_l^2 and the last ones on Y_k^1 vs Y_l^1.

Top-Flop and Threshold betting can handle many cases of binary signals but our setting ans assumptions limit the scope of application. For instance, for political elections, the identical prior assumption is unlikely to hold for any collection of candidates. Our setting also requires that the ratings are still unknown when agents bet. This may pose a problem in cases such as hotel reviews (even if the review is restricted to be binary), when hotels have publicly available ratings. However, the simple bets of this paper could still be used to incentivize honest reporting by test clients in new hotels before opening.

Throughout the paper, we implicitly assumed that the center, offering the bets or organizing the games, is willing to pay up to π for each signal. Often, participation in surveys or experiments is rewarded. What we propose here is to use this reward as prize π, to make agents reveal their signal instead of only rewarding them for providing *any* answer. Our results from the game setting assume that agents cannot communicate. If they could, a full coalition can make sure they get π with probability 1 if K is even and all agents with even items play 1 and all agents with odd items play 0. A way to deter such coalitions is to make the game zero-sum.

5.4.2 Practical implementation and examples

Organizing Top-Flop or Threshold betting on exogenous ratings is easier to implement in practice than the respective game versions. Threshold betting can, for instance, be combined with prediction markets. When people predict the rating of a movie or the results of a song contest, they do not report their own taste but their beliefs about others. Threshold betting, where the rating is defined as the price in the prediction markets for items l and k at a given time, reveals people's own taste (under the assumptions and setting of Section 5.2). A threshold-y bet on prediction market k is a digital option that pays π if the price reaches y. In other words, Top-Flop and Threshold bets can be implemented as derivatives of existing markets.

Let us conclude with two other examples. The director of a company hesitates where to invest in research and development. There is a set $\mathcal{K}$ of possible product features that could be developed. The director would like to know for which feature the consumers would be willing to pay \$100 more. These features do not exist yet and therefore cannot be sold to consumers. Hence, eliciting the willingness-to-pay cannot be incentivized, for instance with the Becker-deGroot-Marschak mechanism (Becker, DeGroot and Marschak, 1964), because it would require actually selling the features. Instead, the director could implement a Top-Flop game among a panel of consumers, organized in K subgroups. Each subgroup of panelists are informed about a feature, and have to bet Top or Flop, not knowing what the other possible innovative features are. A final example of possible application concerns environmental research. It is not always possible to incentivize the elicitation of the willingness-to-pay to save (or the willingness-to-accept for not saving) endangered species. Our simple bets can help there as well by providing subgroups of respondents with information about one (rare) species and ask them whether more people would pay a given amount to save the species they were informed about rather than another random species.

5.5 Conclusion

This paper introduced two methods, Top-Flop and Threshold betting, to elicit private signals. The first part of the paper showed how to transform pre-existing ratings, which may be biased or only partially-informative, into a mechanism incentivizing truthful revelation of signals. An agent betting on the ratings need not fully trust them, but only believe that they are somewhat associated with the signals. In the second part of the paper, the ratings naturally arise from the other agents' betting decisions. In retrospect, our bets, and therefore our mechanisms, look quite simple but they have been overlooked so far, in favor of more complex approaches. The payment rules of Top-Flop and Threshold bets are transparent, with a unique, fixed prize assigned to a well-defined event. We established conditions ensuring that Top-Flop and Threshold betting properly reveal signals. These conditions are milder in terms of individual preferences than typically assumed in the literature, and therefore more likely to be satisfied in practical applications.

Appendix 5.A Proofs for the single-agent setting

Proof of Lemma 5.1

Proof. The posterior cumulative distribution for item l is $F_l^1(y) = 1 - P(Y_l > y \mid t = 1)$. By Bayes rule, we have

$$P(Y_l > y \mid t = 1) = \frac{P(t = 1 \mid Y_l > y)}{P(t = 1)} \times P(Y_l > y). \tag{5.5}$$

By definition, $P(Y_l > y) = 1 - H_l(y)$, and by by Assumption 5.1, $1 - H_l(y) = 1 - H_k(y) = P(Y_k > y)$. Furthermore, Assumption 5.2 states that $P(t = 1 \mid Y_l > y) > P(t = 1 \mid Y_k > y)$ if $y \in \mathcal{S}'$. Hence, we have

$$P(Y_l > y \mid t = 1) > \frac{P(t = 1 \mid Y_k > y)}{P(t = 1)} \times P(Y_k > y) = P(Y_k > y \mid t = 1) \tag{5.6}$$

if $y \in \mathcal{S}'$ and

$$P(Y_l > y \mid t = 1) = P(Y_k > y \mid t = 1) = P(Y_k > y) \tag{5.7}$$

otherwise. As a conclusion, $F_l^1(y) \succ_{SD} F_k^1(y)$ for all $y \in \mathcal{S}$.

We now consider $t = 0$. By definition,

$$P(Y_l > y \mid t = 0) = \frac{P(Y_l > y) - P(Y_l > y \mid t = 1) P(t = 1)}{P(t = 0)} \tag{5.8}$$

and

$$P(Y_k > y \mid t = 0) = \frac{P(Y_k > y) - P(Y_k > y \mid t = 1) P(t = 1)}{P(t = 0)}. \tag{5.9}$$

By Assumption 5.1, $P(Y_l > y) = P(Y_k > y)$ and By Eqs. 5.6 and 5.7, $F_k^1(y) \succ_{SD} F_l^1(y)$ for all $y \in \mathcal{S}$. □

Proof of Lemma 5.2

Proof.

$$\begin{aligned}
P(\{\omega \in \Omega : Y_l(\omega) < Y_k(\omega)\}) &= P\left(\bigcup_{s \in \mathcal{S}} \{\omega \in \Omega : Y_l(\omega) = s\} \cap \{\omega \in \Omega : Y_k(\omega) > s\}\right) \\
&= \sum_{s \in \mathcal{S}} P(\{\omega \in \Omega : Y_l(\omega) = s\} \cap \{\omega \in \Omega : Y_k(\omega) > s\}) \\
&= \sum_{s \in \mathcal{S}} P(\{\omega \in \Omega : Y_l(\omega) = s\}) \times P(\{\omega \in \Omega : Y_k(\omega) > s\}) \\
&= \sum_{s \in \mathcal{S}} P(Y_l = s) \times (1 - H_k(s)).
\end{aligned} \tag{5.10}$$

The second equality comes from events $\{\omega \in \Omega : Y_l(\omega) = s\}$ for any two s being disjoint. Independence (Assumption 5.3) implies the third equality. Because Y_l and Y_k are identically distributed, $P(Y_l = s) = P(Y_k = s)$ and $H_k(s) = H_l(s)$ for all s and therefore, $P(\{\omega \in \Omega : Y_l(\omega) < Y_k(\omega)\}) = P(\{\omega \in \Omega : Y_l(\omega) > Y_k(\omega)\})$. By Assumption 5.4, the agent is indifferent between the Top and the Flop bet.

By Assumption 5.4, the agent would prefer a bet on $\varnothing$ to the Top bet or to the Flop bet if and only if $P(\{\omega \in \Omega : Y_l(\omega) > Y_k(\omega)\}) = 0$ or $P(\{\omega \in \Omega : Y_l(\omega) < Y_k(\omega)\}) = 0$. We have just shown that $P(\{\omega \in \Omega : Y_l(\omega) > Y_k(\omega)\}) = P(\{\omega \in \Omega : Y_l(\omega) < Y_k(\omega)\})$. Hence, the agent would prefer a bet on $\varnothing$ if and only if $P(\{\omega \in \Omega : Y_l(\omega) = Y_k(\omega)\}) = 1$. This implies $P(\{\omega \in \Omega : Y_l(\omega) = Y_k(\omega)\} \mid t = 1) = 1$ and therefore, $F_l^1(y) = F_k^1(y)$. The latter contradicts $F_l^1(y) \succ_{SD} F_k^1(y)$ and according to Lemma 5.1, is therefore incompatible with Assumptions 5.1 and 5.2. As a consequence, under Assumptions 5.1 to 5.4, the agent must strictly prefer any of the bets he is offered to nothing. □

Proof of Theorem 5.1

Proof. Assume $t = 1$.

$$\begin{aligned}
&P(\{\omega \in \Omega : Y_l(\omega) < Y_k(\omega)\} \mid t = 1) \\
=&P\left(\bigcup_{s\in\mathcal{S}}\{\omega \in \Omega : Y_l(\omega) = s\} \cap \{\omega \in \Omega : Y_k(\omega) > s\} \mid t = 1\right) \\
=&\sum_{s\in\mathcal{S}} P(\{\omega \in \Omega : Y_l(\omega) = s\} \cap \{\omega \in \Omega : Y_k(\omega) > s\} \mid t = 1) \\
=&\sum_{s\in\mathcal{S}} P(\{\omega \in \Omega : Y_l(\omega) = s\} \mid t = 1) \times P(\{\omega \in \Omega : Y_k(\omega) > s\} \mid t = 1) \\
=&\sum_{s\in\mathcal{S}} P(Y_l = s \mid t = 1) \times \left(1 - F_k^1(s)\right).
\end{aligned} \tag{5.11}$$

The second equality comes from events $\{\omega \in \Omega : Y_l(\omega) = s\}$ being disjoint for any two s. Conditional independence (Assumption 5.3) implies the third equality.

$$\begin{aligned}
&P(\{\omega \in \Omega : Y_l(\omega) > Y_k(\omega)\} \mid t = 1) \\
=&\sum_{s\in\mathcal{S}} P(Y_k = s \mid t = 1) \times \left(1 - F_l^1(s)\right) \\
>&\sum_{s\in\mathcal{S}} P(Y_k = s \mid t = 1) \times \left(1 - F_k^1(s)\right) \\
\geq&\sum_{s\in\mathcal{S}} P(Y_l = s \mid t = 1) \times \left(1 - F_k^1(s)\right) \\
=&P(\{\omega \in \Omega : Y_l(\omega) < Y_k(\omega)\} \mid t = 1)
\end{aligned} \tag{5.12}$$

The first equality comes from Eq. 5.11 (reversing l and k) and the following inequality from Lemma 5.1 because $F_l^1(s) \succ_{SD} F_k^1(s)$ means that $F_l^1(s) \leq F_k^1(s)$ with a strict inequality for some s. Notice that stochastic dominance also implies that Y_l can be obtained from Y_k by moving probability mass from low values of $\mathcal{S}$ to high values of $\mathcal{S}$. The weights $\left(1 - F_k^1(s)\right)$ are lower for high values of $\mathcal{S}$ than for low values and therefore, replacing Y_k by Y_l decreases the whole sum, which justifies the fourth line of the equation. The final line is obtained from Eq. 5.11.

Together with Assumption 5.4, Eq. 5.12 implies that the agent prefers the Top bet when his signal is $t = 1$. The proof from $t = 0$ is symmetric. □

Proof of Corollary 5.1

Proof. If k is randomly chosen in $\mathcal{K} \setminus \{l\}$, with the random device being independent of all random variables and conditionally independent given T, then the winning probability of the Top and Flop bets does not change and the preferences given in Theorem 5.1 still hold. □

Proof of Lemma 5.3

Proof. Under Assumption 5.1, $H_k(y) = H_l(y) > 0$ for all $y \in \mathcal{S}'$. This, together with Assumption 5.4, gives the result. □

Proof of Theorem 5.2

Proof. From Lemma 5.1, we know $F_l^1(y) \succ_{SD} F_k^1(y)$ and $F_k^0(y) \succ_{SD} F_l^0(y)$ for all $k \neq l$. More precisely, the proof showed $F_l^1(y) < F_k^1(y)$ for all $y \in \mathcal{S}'$, and by symmetry, $F_l^0(y) > F_k^0(y)$. We obtain, for all $y \in \mathcal{S}'$, $P(Y_l > y \mid t = 1) > P(Y_k > y \mid t = 1)$ and $P(Y_l > y \mid t = 0) < P(Y_k > y \mid t = 0)$. Assumption 5.4 then implies the preferences described in the theorem. □

Proof of Corollary 5.2

Proof. If k is randomly chosen in $\mathcal{K} \setminus \{l\}$, with the random device being independent of all random variables and conditionally independent given T, then the winning probability of bets do not change and the preferences given in Theorem 5.1 remain.

If y is drawn from $\mathcal{S}$, either $y \in \mathcal{S}'$ and the strict preferences mentioned in Theorem 5.2 hold, or the events are equally likely and the agent would be indifferent. Hence, before knowing y, the strict preferences mentioned in Theorem 5.2 hold. □

Appendix 5.B Proofs for the game setting

Proof of Theorem 5.3

Proof. Consider $(b_i^0, b_i^1; a^0, a^1)$ with $a_j^0 = 0$ and $a_j^1 = 1$ for all $j \neq i$ and $(b_i^0, b_i^1) \in \mathcal{A}^2$. Hence, in state ω, $\widehat{Y}_{i,k} = \sum_{j \in \mathcal{I}_{i,k}} a_j^{T_j(\omega)} = \sum_{j \in \mathcal{I}_{i,k}} T_j(\omega)$, which implies $\widehat{Y}_{i,k} = Y_{i,k}(\omega)$ for all k, and noticeably for l_i and h_i. Under Assumption 5.5, it is common knowledge that Assumptions 5.1 to 5.4 are satisfied. Therefore, applying Theorem 5.1, it is also common knowledge that agent i strictly prefers $a_i^1 = 1$ to $b_i^1 = 0$ (when b_i^0 is fixed) if $T_i = 1$ and strictly prefers $a_i^0 = 0$ to $b_i^0 = 1$ (when b_i^1 is fixed) if $T_i = 0$. $P_i(T_i = 0 \mid T_i = 1) = P_i(T_i = 1 \mid T_i = 0) = 0$ implies that the agent is indifferent between $a_i^0 = 1$ and $b_i^0 = 0$ (when b_i^1 is fixed) if $T_i = 1$ and between $a_i^1 = 0$ and $b_i^1 = 1$ (when b_i^0 is fixed) if $T_i = 0$. Hence, it is common knowledge that a best response of i to $a_j^0 = 0$ and $a_j^1 = 1$ for all $j \neq i$ is $a_i^0 = 0$ and $a_i^1 = 1$ and therefore, (a^0, a^1) is a Nash equilibrium. It is a strict Nash equilibrium because we showed $(0,1)$ is strictly preferred to $(1,1)$ given $T_i = 0$ and $(0,1)$ is strictly preferred to $(1,0)$ given $T_i = 1$. □

Proof of Corollary 5.3

Proof. Note that the strategy profiles with $b_i^0 = b_i^1 = 0$ for all i gives payment 0 to everyone. The same is true for $b_i^0 = b_i^1 = 1$. By contrast, the equilibrium in Theorem 5.3 is strict, which would not be possible if the payment was 0. □

Proof of Theorem 5.4

Proof. The proof is similar to that of Theorem 5.3, simply using Theorem 5.2 instead of Theorem 5.1. □

Proof of Corollary 5.4

Proof. The proof is similar to that of Corollary 5.2. □

Chapter 6

Conclusion

This dissertation develops new elicitation methods for individual preferences and private information and experimentally tests them. Chapter 2 and Chapter 3 extend the classical revealed preference approach to elicit individuals' non-standard preferences. Chapter 4 and Chapter 5 study the incentive alignment in truth-telling mechanisms. In particular, Chapter 2 recovers individuals' preferences over wealth distributions behind the veil of ignorance. Chapter 3 elicits preferences over useless and useful information sources (tests). Chapter 4 tests whether and how Bayesian markets induce truth-telling. Chapter 5 introduces two simple methods, Top-Flop and Threshold betting, to elicit private signals.

Chapter 2 studies what people like regarding the wealth distribution of a society. We develop a new instrument and elicit individual preferences over the trade-off between the aggregate wealth and the distributional equity. Our instrument is robust since it provides individuals a rich set of trade-off problems. Our instrument is also clean. It extracts self-serving motives by imposing individuals behind the veil of ignorance, and it disentangles individual risk preferences by pairing each VoI problem with a risk problem. We test the veracity of the instrument using a within-subject laboratory experiment. We find clusters of equity preferring, efficiency preferring, and egoist individuals through reduced form, revealed preference, and structural estimation analyses. Such a preference heterogeneity is hidden at the aggregation level. Moreover, we find that individuals' preferences for equity and efficiency behind the veil of ignorance are not correlated with their risk preferences. These findings have important implications for the aggregating problem in welfare economics and the preference-based redistribution policy designs.

Chapter 3 studies what people like for information sources. I construct trade-offs between different motives relevant for how individuals choose and evaluate quack and expert tests. Using a laboratory experiment, I elicit individual preferences over tests. The experiment is user-friendly. It translates a linear budget set of tests into a graphic coloring task. It also provides clean diagnoses for different mechanisms contributing to quack choices. I find people fail to distinguish experts and quacks. They frequently select quacks even though many decision-enhancing expert tests are also in their choice set. They are overpaying for quacks but accurately paying for experts. People either choose the most useful expert or the most distant quack. These findings are not explained by standard behavioral biases, including belief-updating bias, sub-optimal action choices, or intrinsic preferences over test characteristics. Instead, they confirm a universal bias in reasoning when a test is useful. That is, people fail to reason contingently on how different test structures interact with the decision problem at hand. Such failures of contingent reasoning provide new implications for how people seek out information sources, how data sellers provide information products, and the general information design problems. They also suggest that we need new de-biasing interventions to make people reason contingently.

Chapter 4 experimentally tests the validity of Bayesian markets in inducing private signals and examines the belief requirement for the truth-telling equilibrium. I construct Bayesian markets with different distortions in participants' beliefs over their opponents' truthfulness. I find Bayesian markets are significantly less effective when participants suspect their opponents lie. Further investigations of participants' transaction decisions reveal how different types of traders in the market contribute to updating biases, resulting in market bubbles and underperformance of the mechanism. These findings emphasize the potential importance of market stabilizer schemes for the practical design of Bayesian markets. They also imply some general problems for other market-based elicitation methods. Market forces may not eliminate nonstandard behaviors. Instead, they provide interactions among heterogeneous participants. A small fraction of speculative buyers in the market exacerbates behavioral biases and further amplifies through the aggregate price.

Chapter 5 proposes two simple betting mechanisms to extract private information. Both mechanisms relax some heavy assumptions in the literature and are practically simple. They offer agents bets on the relative ratings between two items, one about which they have a private signal and the other one about which they do not. When the ratings are exogenous, we characterize the key conditions (unique prior and comparative informativeness) for the agents to reveal their private signals through their bet choices. When the ratings are endogenously determined by the actions of other agents, we establish microfoundations of the ratings in a game setting and still obtain that bet choices reveal private signals. Using a reference item (instead of a reference peer in standard mechanisms), our mechanisms alleviate the concern of Keynesian beauty-contest type of herding. Individuals act upon their private signals, not what they think others will think.

Rabin (2002) argued that in the "post-anomalies" wave of behavioral economics, we are now on to the task of integrating behavioral insights into the traditional economics. The biases and anomalies, assumption deviations, and alternative conceptualizations should be studied using standard economic methods; tested using standard econometric techniques; and judged by standard scientific criteria such as parsimony, prediction, generality, insight. This dissertation is such an attempt. It employs standard tools in neoclassical economics to disentangle the confounding motives and systematic biases behind people's preferences, beliefs, and information processing. The dissertation improves our understanding of what people like and what people think in new contexts.

Summary

This dissertation develops new elicitation methods for preferences and private information and experimentally test them. Chapter 2 recovers individual preferences over the trade-off between the aggregate wealth and the distributional equity behind the veil of ignorance and investigates their relationships with risk preferences. Chapter 3 elicits individual preferences over expert and quack tests and identifies how failures of contingent reasoning contribute to the choices of quacks. Chapter 4 tests the validity of the Bayesian market in eliciting private information and examines the role of belief disturbances. Chapter 5 relaxes standard assumptions and proposes two simple betting mechanisms to extract private information.

In the "post-anomalies" wave of behavioral economics, we are now on to the task of integrating behavioral insights into the traditional economics. Rabin (2002) argued that the biases and anomalies, assumption deviations, and alternative conceptualizations should be studied using standard economic methods; tested using standard econometric techniques; and judged by standard scientific criteria such as parsimony, prediction, generality, insight. This dissertation is such an attempt. It employs standard tools in neoclassical economics to disentangle the confounding motives and systematic biases behind people's preferences, beliefs, and information processing. The dissertation improves our understanding of what people like and what people think in new contexts.

Samenvatting

Samenvattend, dit proefschrift ontwikkelt nieuwe methoden om voorkeuren en privéinformatie te achterhalen en test deze methoden door middel van experimenten. Hoofdstuk 2 gebruikt de zo genaamde “veil of ignorance” om persoonlijke voorkeuren met betrekking tot de afweging tussen het maximaliseren van totale rijkdom en het zorgen voor een gelijke verdeling te detecteren en onderzoekt hoe dat relateert aan risico voorkeuren. Hoofdstuk 3 achterhaald individuele voorkeuren met betrekking tussen “experts en kwakzalver”-testen en identificeert hoe fouten in voorwaardelijke redenatie bijdraagt aan het verkiezen van kwakzalvers boven experts. Hoofdstuk 4 test de validiteit van de Bayesiaanse markt bij het meten van privéinformatie en onderzoekt de rol van verstoringen van overtuigingen. Hoofdstuk 5 versoepelt de standaard aannames en stelt twee simpele bied-mechanismen voor om privéinformatie te extraheren.

In de ′post-abnormaliteiten′-golf in de gedragseconomie hebben we de taak om gedragsinzichten te integreren in de traditionele economie. Rabin (2002) beschrijft dat de vooroordelen en afwijkingen, afwijkingen in veronderstellingen en alternatieve conceptualisaties moeten worden bestudeerd met behulp van standaard economische methoden; getest met standaard econometrische technieken; en beoordeeld aan de hand van standaard wetenschappelijke criteria zoals spaarzaamheid, voorspelbaarheid, algemeenheid en inzichtelijkheid. Dit proefschrift is hierin een poging. Met het gebruik van de standaard tools uit de neoklassieke economie worden de schijnbaar willekeurige motieven en systematische vooroordelen achter voorkeuren, overtuigingen en de manier van informatieverwerking getracht te ontrafelen. Dit proefschrift schijnt nieuw licht op ons begrip in hoe mensen denken en hoe voorkeuren ontstaan.

Bibliography

Acemoglu, Daron, Simon Johnson and James A. Robinson. 2002. "Reversal of fortune: Geography and institutions in the making of the modern world income distribution." *Quarterly Journal of Economics* 117(4):1231–1294.

Afriat, Sydney N. 1967. "The construction of utility functions from expenditure data." *International Economic Review* 8(1):67–77.

Akaike, Hirotogu. 1998. Information theory and an extension of the maximum likelihood principle. In *Selected Papers of Hirotugu Akaike*. Springer pp. 199–213.

Alesina, Alberto and Dani Rodrik. 1994. "Distributive politics and economic growth." *Quarterly Journal of Economics* 109(2):465–490.

Alesina, Alberto and Roberto Perotti. 1996. "Income distribution, political instability, and investment." *European Economic Review* 40(6):1203–1228.

Ambuehl, Sandro and Shengwu Li. 2018. "Belief updating and the demand for information." *Games and Economic Behavior* 109:21–39.

Amiel, Yoram, Frank A Cowell and Wulf Gaertner. 2009. "To be or not to be involved: A questionnaire-experimental view on Harsanyi's utilitarian ethics." *Social Choice and Welfare* 32(2):299–316.

Amiel, Yoram and Frank Cowell. 2000. Attitudes to risk and inequality: A new twist on the transfer principle. Technical report Distributional Analysis Discussion Paper.

Andersson, Ola, Håkan J Holm, Jean-Robert Tyran and Erik Wengström. 2014. "Deciding for others reduces loss aversion." *Management Science* 62(1):29–36.

Andreoni, James and Charles Sprenger. 2012. "Estimating time preferences from convex budgets." *American Economic Review* 102(7):3333–56.

Andreoni, James and John Miller. 2002. "Giving according to GARP: An experimental test of the consistency of preferences for altruism." *Econometrica* 70(2):737–753.

Arrow, Kenneth J, Robert Forsythe, Michael Gorham, Robert Hahn, Robin Hanson, John O Ledyard, Saul Levmore, Robert Litan, Paul Milgrom, Forrest D Nelson et al. 2008. "The promise of prediction markets." *Science* 320:877–878.

Azrieli, Yaron, Christopher P. Chambers and Paul J. Healy. forthcoming. "Incentives in experiments: A theoretical analysis." *Journal of Political Economy* .

Baillon, Aurélien. 2016. "A market to read minds." *Working Paper* .

Baillon, Aurélien. 2017. "Bayesian markets to elicit private information." *Proceedings of the National Academy of Sciences* 114(30):7958–7962.

Baillon, Aurélien and Yan Xu. 2019. "Simple bets to elicit private signals." *Working Paper* .

Barberis, Nicholas C. 2013. "Thirty years of prospect theory in economics: A review and assessment." *Journal of Economic Perspectives* 27(1):173–96.

Baron, David P and Roger B Myerson. 1982. "Regulating a monopolist with unknown costs." *Econometrica* pp. 911–930.

Becker, Gordon M, Morris H DeGroot and Jacob Marschak. 1964. "Measuring utility by a single-response sequential method." *Behavioral Science* 9(3):226–232.

Becker, Nicole, Kirsten Häger and Jan Heufer. 2013. "Revealed notions of distributive justice II: Experimental analysis." *Ruhr Economic Papers* #444, TU Dortmund University, Discussion Paper. working paper.

Benabou, Roland. 2000. "Unequal societies: Income distribution and the social contract." *American Economic Review* 90(1):96–129.

Benjamin, Daniel J. 2019. Errors in probabilistic reasoning and judgment biases. In *Handbook of Behavioral Economics: Applications and Foundations 1*. Vol. 2 Elsevier pp. 69–186.

Berg, Joyce E, Forrest D Nelson and Thomas A Rietz. 2008. "Prediction market accuracy in the long run." *International Journal of Forecasting* 24(2):285–300.

Berg, Joyce, Robert Forsythe, Forrest Nelson and Thomas Rietz. 2008. "Results from a dozen years of election futures markets research." *Handbook of Experimental Economics Results* 1:742–751.

Bergemann, D and S Morris. 2009*a*. "Virtual Robust Implementation." *Theoretical Economics* 4:45–88.

Bergemann, Dirk and Alessandro Bonatti. 2019. "Markets for information: An introduction." *Annual Review of Economics* 11:85–107.

Bergemann, Dirk, Alessandro Bonatti and Alex Smolin. 2018. "The design and price of information." *American Economic Review* 108(1):1–48.

Bergemann, Dirk and Stephen Morris. 2005. "Robust mechanism design." *Econometrica* pp. 1771–1813.

Bergemann, Dirk and Stephen Morris. 2009*b*. "Robust implementation in direct mechanisms." *The Review of Economic Studies* 76(4):1175–1204.

Bergemann, Dirk and Stephen Morris. 2019. "Information design: A unified perspective." *Journal of Economic Literature* 57(1):44–95.

Bernasconi, Michele. 2002. "How should income be divided? Questionnaire evidence from the theory of "Impartial preferences"." *Journal of Economics* 9(1):163–195.

Blackwell, David. 1951. Comparison of Experiments. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*. University of California Press pp. 93–102.

Bordalo, Pedro, Nicola Gennaioli and Andrei Shleifer. 2012. "Salience theory of choice under risk." *The Quarterly Journal of Economics* 127(3):1243–1285.

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

Bronars, Stephen G. 1987. "The Power of nonparametric tests of preference maximization." *Econometrica* 55(3):693–698.

Cabrales, Antonio, Olivier Gossner and Roberto Serrano. 2013. "Entropy and the value of information for investors." *American Economic Review* 103(1):360–77.

Camerer, Colin. 1999. "Behavioral economics: Reunifying psychology and economics." *Proceedings of the National Academy of Sciences* 96(19):10575–10577.

Camerer, Colin F, Anna Dreber, Eskil Forsell, Teck-Hua Ho, Jürgen Huber, Magnus Johannesson, Michael Kirchler, Johan Almenberg, Adam Altmejd, Taizan Chan et al. 2016. "Evaluating replicability of laboratory experiments in economics." *Science* 351(6280):1433–1436.

Camerer, Colin F, Anna Dreber, Felix Holzmeister, Teck-Hua Ho, Jürgen Huber, Magnus Johannesson, Michael Kirchler, Gideon Nave, Brian A Nosek, Thomas Pfeiffer et al. 2018. "Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015." *Nature Human Behaviour* 2(9):637–644.

Camerer, Colin F and George Loewenstein. 2003. "Behavioral economics: Past, present, future.".

Camerer, Colin F and Richard H Thaler. 1995. "Anomalies: Ultimatums, dictators and manners." *Journal of Economic perspectives* 9(2):209–219.

Cason, Timothy N and Charles R Plott. 2014. "Misconceptions and game form recognition: Challenges to theories of revealed preference and framing." *Journal of Political Economy* 122(6):1235–1270.

Chan, David C, Matthew Gentzkow and Chuan Yu. 2019. "Selection with variation in diagnostic skill: Evidence from radiologists." *Working Paper* .

Charness, Gary and Matthew Rabin. 2002. "Understanding social preferences with simple tests." *Quarterly Journal of Economics* 117:817–869.

Charness, Gary, Ryan Oprea and Sevgi Yuksel. 2018. "How do people choose between biased information sources? Evidence from a laboratory experiment." *Working Paper* .

Chen, Yan. 2008. "Incentive-compatible mechanisms for pure public goods: A survey of experimental research." *Handbook of Experimental Economics Results* 1:625–643.

Choi, Syngjoo, Raymond Fisman, Douglas Gale and Shachar Kariv. 2007*a*. "Consistency and heterogeneity of individual behavior under uncertainty." *American Economic Review* 97(5):1921–1938.

Choi, Syngjoo, Raymond Fisman, Douglas Gale and Shachar Kariv. 2007*b*. "Revealing preferences graphically: An old method gets a mew tool kit." *American Economic Review* 97(2):153–158.

Choi, Syngjoo, Shachar Kariv, Wieland Müller and Dan Silverman. 2014. "Who is (more) rational?" *American Economic Review* 104(6):1518–1550.

Costa-Gomes, Miguel A and Georg Weizsäcker. 2008. "Stated beliefs and play in normal-form games." *The Review of Economic Studies* 75(3):729–762.

Cox, James C. 1997. "On testing the utility hypothesis." *The Economic Journal* 107(443):1054–1078.

Cox, James C. 2004. "How to identify trust and reciprocity." *Games and Economic Behavior* 46(2):260–281.

Cox, James, Vjollca Sadiraj and Ulrich Schmidt. 2015. "Paradoxes and mechanism for choice under risk." *Experimental Economics* 18(2):215–250.

Crémer, Jacques and Richard P McLean. 1988. "Full extraction of the surplus in Bayesian and dominant strategy auctions." *Econometrica* pp. 1247–1257.

Cvitanić, Jakša, Dražen Prelec, Blake Riley, Benjamin Tereick et al. 2019. "Honesty via choice-matching." *American Economic Review: Insights* 1(2):179–92.

Dasgupta, Anirban and Arpita Ghosh. 2013. Crowdsourced judgement elicitation with endogenous proficiency. In *Proceedings of the 22nd International Conference on World Wide Web*. WWW '13 New York, NY, USA: ACM pp. 319–330.

DellaVigna, Stefano. 2009. "Psychology and economics: Evidence from the field." *Journal of Economic literature* 47(2):315–72.

DellaVigna, Stefano and Matthew Gentzkow. 2010. "Persuasion: Empirical evidence." *Annual Review of Economics* 2(1):643–669.

Dertwinkel-Kalt, Markus and Mats Köster. 2019. "Salience and skewness preferences." *Journal of the European Economic Association* .

Dillenberger, David. 2010. "Preferences for one-shot resolution of uncertainty and Allais-type behavior." *Econometrica* 78(6):1973–2004.

Dreber, Anna, Thomas Pfeiffer, Johan Almenberg, Siri Isaksson, Brad Wilson, Yiling Chen, Brian A Nosek and Magnus Johannesson. 2015. "Using prediction markets to estimate the reproducibility of scientific research." *Proceedings of the National Academy of Sciences* 112(50):15343–15347.

Ely, Jeffrey, Alexander Frankel and Emir Kamenica. 2015. "Suspense and surprise." *Journal of Political Economy* 123(1):215–260.

Engelmann, Dirk and Martin Strobel. 2004. "Inequality aversion, efficiency, and maximin preferences in simple distribution experiments." *American Economic Review* 94(4):857–869.

Enke, Benjamin and Thomas Graeber. 2019. "Cognitive uncertainty." *Working Paper* .

Epstein, Larry G and Stanley E Zin. 1989. "Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework." *Econometrica* 57(4):937–969.

Eső, Péter and Balazs Szentes. 2007*a*. "Optimal information disclosure in auctions and the handicap auction." *The Review of Economic Studies* 74(3):705–731.

Eső, Péter and Balázs Szentes. 2007*b*. "The price of advice." *The Rand Journal of Economics* 38(4):863–880.

Esponda, Ignacio and Emanuel Vespa. 2014. "Hypothetical thinking and information extraction in the laboratory." *American Economic Journal: Microeconomics* 6(4):180–202.

Esponda, Ignacio and Emanuel Vespa. 2019. "Contingent thinking and the sure-thing principle: Revisiting classic anomalies in the laboratory." *Working Paper* .

Falk, Armin and Florian Zimmermann. 2016. "Beliefs and utility: Experimental evidence on preferences for information." *Working Paper* .

Fama, Eugene F. 1970. "Efficient capital markets: A review of theory and empirical work." *The Journal of Finance* 25(2):383–417.

Fisman, Raymond, Pamela Jakiela and Shachar Kariv. 2017. "Distributional preferences and political behavior." *Journal of Public Economics* 155:1–10.

Fisman, Raymond, Pamela Jakiela, Shachar Kariv and Daniel Markovits. 2015. "The distributional preferences of an elite." *Science* 349(6254).

Fisman, Raymond, Shachar Kariv and Daniel Markovits. 2007. "Individual preferences for giving." *American Economic Review* 97(5):1858–1876.

Frederick, Shane. 2005. "Cognitive reflection and decision making." *Journal of Economic Perspectives* 19(4):25–42.

Frignani, Nicola and Giovanni Ponti. 2012. "Risk versus social preferences under the veil of ignorance." *Economics Letters* 116(2):143–146.

Füllbrunn, Sascha C and Wolfgang J Luhan. 2017. "Decision making for others: The case of loss aversion." *Economics Letters* 161:154–156.

Füllbrunn, Sascha and Wolfgang Luhan. 2015. "Am I my peer's keeper? Social responsibility in financial decision making." *Ruhr Economic Papers #551* .

Ganguly, Ananda and Joshua Tasoff. 2017. "Fantasy and dread: The demand for information and the consumption utility of the future." *Management Science* 63(12):4037–4060.

Gao, Xi Alice, Andrew Mao, Yiling Chen and Ryan Prescott Adams. 2014. Trick or treat: Putting peer prediction to the test. In *Proceedings of the 15th ACM Conference on Economics and Computation*. ACM pp. 507–524.

Gigerenzer, Gerd Ed, Ralph Ed Hertwig and Thorsten Ed Pachur. 2011. *Heuristics: The foundations of adaptive behavior*. Oxford University Press.

Gigerenzer, Gerd and Ulrich Hoffrage. 1995. "How to improve Bayesian reasoning without instruction: Frequency formats." *Psychological Review* 102(4):684.

Gneiting, Tilmann and Adrian E Raftery. 2007. "Strictly proper scoring rules, prediction, and estimation." *Journal of the American statistical Association* 102(477):359–378.

Gneiting, Tilmann and Matthias Katzfuss. 2014. "Probabilistic forecasting." *Annual Review of Statistics and Its Application* 1:125–151.

Gonzalez, Richard and George Wu. 1999. "On the shape of the probability weighting function." *Cognitive Psychology* 38(1):129–166.

Good, Isidore Jacob. 1950. "Probability and the weighing of evidence.".

Grant, Simon, Atsushi Kajii and Ben Polak. 1998. "Intrinsic preference for information." *Journal of Economic Theory* 83(2):233–259.

Green, Jerry and Nancy Stokey. 1978. *Two representations of information structures and their comparisons*. Institute for Mathematical Studies in the Social Sciences.

Greiner, Ben. 2004. The Online recruitment system ORSEE 2.0 – A Guide for the organization of experiments in economics. Technical report University of Cologne. Technical Report 10, Working Paper Series in Economics.

Grether, David M. 1980. "Bayes rule as a descriptive model: The representativeness heuristic." *The Quarterly Journal of Economics* 95(3):537–557.

Gul, Faruk. 1991. "A theory of disappointment aversion." *Econometrica* 59(3):667–686.

Hadar, Josef and William R. Russell. 1969. "Rules for ordering uncertain prospects." *American Economic Review* 59(1):25–33.

Hanley, James A and Barbara J McNeil. 1982. "The meaning and use of the area under a receiver operating characteristic (ROC) curve." *Radiology* 143(1):29–36.

Harbaugh, William T, Kate Krause and Timothy R Berry. 2001. "GARP for kids: On the development of rational choice behavior." *American Economic Review* 91(5):1539–1545.

Harsanyi, John C. 1953. "Cardinal utility in welfare economics and in the theory of risk-taking." *Journal of Political Economy* 61(5):434–435.

Harsanyi, John C. 1968. "Games with incomplete information played by "Bayesian" players part II. Bayesian equilibrium points." *Management Science* 14(5):320–334.

Harsanyi, John C. 1976. Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. In *Essays on Ethics, Social Behavior, and Scientific Explanation*. Springer pp. 6–23.

Harstad, Ronald M. 2000. "Dominant strategy adoption and bidders' experience with pricing rules." *Experimental Economics* 3(3):261–280.

Hayek, Friedrich August. 1945. "The use of knowledge in society." *American Economic Review* 35(4):519–530.

Heufer, Jan. 2013. "Testing revealed preferences for homotheticity with two-good experiments." *Experimental Economics* 16(1):114–124.

Heufer, Jan. 2014. "Nonparametric comparative revealed risk aversion." *Journal of Economic Theory* 153:569–616.

Heufer, Jan, Jason Shachat and Yan Xu. 2019. "Measuring tastes for equity and aggregate wealth behind the veil of ignorance." *Working Paper* .

Heufer, Jan and Per Hjertstrand. 2017. "Homothetic efficiency: Theory and applications." *Journal of Business & Economic Statistics* .

Hirshleifer, Jack. 1971. "The private and social value of information and the reward to inventive activity." *American Economic Review* 61(4):561–74.

Hong, Hao, Jianfeng Ding and Yang Yao. 2015. "Individual social welfare preferences: An experimental study." *Journal of Behavioral and Experimental Economics* 57:89–97.

John, Leslie K, George Loewenstein and Dražen Prelec. 2012. "Measuring the prevalence of questionable research practices with incentives for truth telling." *Psychological Science* pp. 524–532.

Joseph, Wright. 2008. "Do authoritarian institutions constrain? How legislatures affect economic growth and investment." *American Journal of Political Science* 52(2):322–343.

Jurca, Radu and Boi Faltings. 2006. Minimum payments that reward honest reputation feedback. In *Proceedings of the 7th ACM Conference on Electronic Commerce*. ACM pp. 190–199.

Jurca, Radu and Boi Faltings. 2009. "Mechanisms for making crowds truthful." *Journal of Artificial Intelligence Research* 34(1):209–253.

Kahneman, Daniel. 2003*a*. "Maps of bounded rationality: Psychology for behavioral economics." *American Economic Review* 93(5):1449–1475.

Kahneman, Daniel. 2003*b*. "A perspective on judgment and choice: Mapping bounded rationality." *American Psychologist* 58(9):697.

Kahneman, Daniel and Amos Tversky. 1979. "Prospect theory: An analysis of decision under risk." *Econometrica* 47(2):263.

Kahneman, Daniel, Jack L Knetsch and Richard H Thaler. 1991. "Anomalies: The endowment effect, loss aversion, and status quo bias." *Journal of Economic Perspectives* 5(1):193–206.

Kamenica, Emir. 2019. "Bayesian persuasion and information design." *Annual Review of Economics* 11:249–272.

Kamenica, Emir and Matthew Gentzkow. 2011. "Bayesian persuasion." *American Economic Review* 101(6):2590–2615.

Kariv, Shachar and William Zame. 2014. "Character and candidates: A view from decision theory.".

Kolmogoroff, Andrej N. 1933. *Grundbegriffe der wahrscheinlichkeitsrechnung*. Springer, Berlin. Translated into English by Nathan Morrison (1950) *Foundations of the Theory of Probability*, Chelsea, New York.

Kreps, David M and Evan L Porteus. 1978. "Temporal resolution of uncertainty and dynamic choice theory." *Econometrica* pp. 185–200.

Kuziemko, Ilyana, Michael I Norton, Emmanuel Saez and Stefanie Stantcheva. 2015. "How elastic are preferences for redistribution? Evidence from randomized survey experiments." *American Economic Review* 105(4):1478–1508.

Lara, Michel De and Olivier Gossner. 2020. "Payoffs-beliefs duality and the value of information." *SIAM Journal on Optimization* 30(1):464–489.

Lerner, Jennifer S, Ye Li, Piercarlo Valdesolo and Karim S Kassam. 2015. "Emotion and decision making." *Annual Review of Psychology* 66.

Levitt, Steven D and John A List. 2007. "What do laboratory experiments measuring social preferences reveal about the real world?" *Journal of Economic perspectives* 21(2):153–174.

Li, Hao and Xianwen Shi. 2017. "Discriminatory information disclosure." *American Economic Review* 107(11):3363–85.

Li, Jing, William H. Dow and Shachar Kariv. 2017. "Social preferences of future physicians." *Proceedings of the National Academy of Sciences* 114(48):E10291–E10300.

Lizzeri, Alessandro. 1999. "Information revelation and certification intermediaries." *The Rand Journal of Economics* pp. 214–231.

Loewenstein, George. 1987. "Anticipation and the valuation of delayed consumption." *The Economic Journal* 97(387):666–684.

Machina, Mark J. and David Schmeidler. 1992. "A more robust definition of subjective probability." *Econometrica* 60(4):745–780.

Manski, Charles F. 2006. "Interpreting the predictions of prediction markets." *Economics Letters* 91(3):425–429.

Martínez-Marquina, Alejandro, Muriel Niederle and Emanuel Vespa. 2019. "Failures in contingent reasoning: The role of uncertainty." *American Economic Review* 109(10):3437–74.

Masatlioglu, Yusufcan, A Yesim Orhun and Collin Raymond. 2017. "Intrinsic information preferences and skewness." *Working Paper* .

Maskin, Eric. 1999. "Nash equilibrium and welfare optimality." *The Review of Economic Studies* 66(1):23–38.

Maskin, Eric and John Riley. 1984. "Monopoly with incomplete information." *The Rand Journal of Economics* 15(2):171–196.

Michelbach, Philip A, John T Scott, Richard E Matland and Brian H Bornstein. 2003. "Doing Rawls justice: An experimental study of income distribution norms." *American Journal of Political Science* 47(3):523–539.

Miller, Nolan, Paul Resnick and Richard Zeckhauser. 2005. "Eliciting informative feedback: The peer-prediction method." *Management Science* 51(9):1359–1373.

Mirrlees, James A. 1971. "An exploration in the theory of optimum income taxation." *The Review of Economic Studies* 38(2):175–208.

Montanari, Giovanni and Salvatore Nunnari. 2019. "Audi alteram partem: An experiment on selective exposure to information." *Working Paper* .

Moore, Don A and Paul J Healy. 2008. "The trouble with overconfidence." *Psychological Review* 115(2):502.

Moré, Jorge J. 1978. The Levenberg-Marquardt algorithm: Implementation and theory. In *Numerical analysis*. Springer pp. 105–116.

Mullainathan, Sendhil and Richard H Thaler. 2000. Behavioral Economics. Technical report National Bureau of Economic Research.

Myerson, Roger B. 1986. "Multistage games with communication." *Econometrica: Journal of the Econometric Society* pp. 323–358.

Myerson, Roger B. 1991. *Game theory: Analysis of conflict*. Harvard University Press.

Nielsen, Kirby. 2018. "Preferences for the resolution of uncertainty and the timing of information." *Working Paper* .

Offerman, Theo, Joep Sonnemans, Gijs Van de Kuilen and Peter P Wakker. 2009. "A truth serum for non-bayesians: Correcting proper scoring rules for risk attitudes." *The Review of Economic Studies* 76(4):1461–1489.

Ollár, Mariann and Antonio Penta. 2017. "Full implementation and belief restrictions." *American Economic Review* 107(8):2243–77.

Ollár, Mariann, Antonio Penta et al. 2019. "Implementation via transfers with identical but unknown distributions." *Working Paper* .

Osborne, Martin J and Ariel Rubinstein. 1994. *A course in game theory*. MIT press.

Pahlke, Julius, Sebastian Strasser and Ferdinand M Vieider. 2015. "Responsibility effects in decision making under risk." *Journal of Risk and Uncertainty* 51(2):125–146.

Parkes, David C and Jens Witkowski. 2012. A robust Bayesian truth serum for small populations. In *Proceedings of the 26th AAAI Conference on Artificial Intelligence*. pp. 1492–1498.

Prelec, Dražen. 1998. "The probability weighting function." *Econometrica* pp. 497–527.

Prelec, Dražen. 2004. "A Bayesian truth serum for subjective data." *Science* 306(5695):462–466.

Rabin, Matthew. 1998. "Psychology and economics." *Journal of Economic Literature* 36(1):11–46.

Rabin, Matthew. 2002. "A perspective on psychology and economics." *European economic review* 46(4-5):657–685.

Rabin, Matthew and Richard H Thaler. 2001. "Anomalies: Risk aversion." *Journal of Economic perspectives* 15(1):219–232.

Radanovic, Goran and Boi Faltings. 2013. A robust bayesian truth serum for non-binary signals. In *Proceedings of the 27th AAAI Conference on Artificial Intelligence*. pp. 833–839.

Radanovic, Goran and Boi Faltings. 2014. Incentives for truthful information elicitation of continuous signals. In *Proceedings of the 28th AAAI Conference on Artificial Intelligence*. pp. 770–776.

Rawls, John. 1958. "Justice as fairness." *The Philosophical Review* 67(2):164–194.

Rodrik, Dani and Romain Wacziarg. 2005. "Do democratic transitions produce bad economic outcomes?" *American Economic Review* 95(2):50–55.

Saez, Emmanuel and Stefanie Stantcheva. 2016. "Generalized social marginal welfare weights for optimal tax theory." *American Economic Review* 106(1):24–45.

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

Savage, Leonard J. 1972. *The foundations of statistics.* Courier Corporation.

Schildberg-Hörisch, Hannah. 2010. "Is the veil of ignorance only a concept about risk? An experiment." *Journal of Public Economics* 94(11-12):1062–1066.

Schlag, Karl H, James Tremewan and Joel J Van der Weele. 2015. "A penny for your thoughts: A survey of methods for eliciting beliefs." *Experimental Economics* 18(3):457–490.

Schotter, Andrew and Isabel Trevino. 2014. "Belief elicitation in the laboratory." *Annual Review of Economics* 6(1):103–128.

Shannon, Claude E. 1948. "A mathematical theory of communication." *Bell System Technical Journal* 27(3):379–423.

Shaw, Aaron D, John J Horton and Daniel L Chen. 2011. Designing incentives for inexpert human raters. In *Proceedings of the ACM 2011 Conference on Computer Supported Cooperative Work.* ACM pp. 275–284.

Shnayder, Victor, Arpit Agarwal, Rafael Frongillo and David C Parkes. 2016. Informed truthfulness in multi-task peer prediction. In *Proceedings of the 2016 ACM Conference on Economics and Computation.* EC '16 ACM New York, NY, USA: pp. 179–196.

Sippel, Reinhard. 1997. "An experiment on the pure theory of consumer's behaviour." *The Economic Journal* 107(444):1431–1444.

Starmer, Chris. 2000. "Developments in non-expected utility theory: The hunt for a descriptive theory of choice under risk." *Journal of economic literature* 38(2):332–382.

Sujoy, Chakravarty, Harrison Glenn W., Haruvy Ernan E. and Rutström E. Elisabet. 2011. "Are you risk averse over other people's money?" *Southern Economic Journal* 77(4):901–913.

Thaler, Richard H and LJ Ganser. 2015. *Misbehaving: The making of behavioral economics.* WW Norton New York.

Traub, Stefan, Christian Seidl and Ulrich Schmidt. 2009. "An experimental study on individual Choice, social Welfare, and social preferences." *European Economic Review* 53(4):385–400.

Traub, Stefan, Christian Seidl, Ulrich Schmidt and Maria Vittoria Levati. 2005. "Friedman, Harsanyi, Rawls, Boulding–or somebody else? An experimental investigation of distributive justice." *Social Choice and Welfare* 24(2):283–309.

Trautmann, Stefan T and Gijs van de Kuilen. 2018. "Higher order risk attitudes: A review of experimental evidence." *European Economic Review* 103:108–124.

Tsakas, Elias. 2020. "Robust scoring rules." *Theoretical Economics* 15(3):955–987.

Tversky, Amos and Daniel Kahneman. 1979. "Prospect theory: An analysis of decision under risk." *Econometrica* 47(2):263–291.

Tversky, Amos and Daniel Kahneman. 1992. "Advances in prospect theory: Cumulative representation of uncertainty." *Journal of Risk and Uncertainty* 5(4):297–323.

Tversky, Amos and Eldar Shafir. 1992. "Choice under conflict: The dynamics of deferred decision." *Psychological science* 3(6):358–361.

Tversky, Amos and Peter Wakker. 1995. "Risk attitudes and decision weights." *Econometrica* pp. 1255–1280.

Tziralis, George and Ilias Tatsiopoulos. 2007. "Prediction markets: An extended literature review." *The Journal of Prediction Markets* 1(1):75–91.

van Bruggen, Paul and Jan Heufer. 2017. "Afriat in the lab." *Journal of Economic Theory* 169:546–550.

Varian, Hal R. 1982. "The nonparametric approach to demand analysis." *Econometrica* 50(4):945–972.

Varian, Hal R. 1983. "Non-parametric tests of consumer behaviour." *The Review of Economic Studies* 50(1):99–110.

Varian, Hal R. 1993. "Goodness-of-fit for revealed preference tests." *Unpublished manuscript* .

Vieider, Ferdinand M, Clara Villegas-Palacio, Peter Martinsson and Milagros Mejía. 2016. "Risk taking for oneself and others: A structural model approach." *Economic Inquiry* 54(2):879–894.

Wakker, Peter P. 2008. "Explaining the characteristics of the power (CRRA) utility family." *Health Economics* 17(12):1329–1344.

Wakker, Peter P. 2010. *Prospect theory: For risk and ambiguity*. Cambridge University Press.

Wason, Peter C. 1968. "Reasoning about a rule." *Quarterly Journal of Experimental Psychology* 20(3):273–281.

Weaver, Ray and Dražen Prelec. 2013. "Creating truth-telling incentives with the Bayesian truth serum." *Journal of Marketing Research* 50(3):289–302.

Wilson, R.B. 1987. Game-theoretic analyses of trading processes. In *Advances in Economic Theory: Fifth World Congress*, ed. T.F. Bewley. Cambridge: Cambridge University Press chapter 2, pp. 33–70.

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

Winkler, Robert L and Allan H Murphy. 1968. ""Good" probability assessors." *Journal of applied Meteorology* 7(5):751–758.

Witkowski, Jens and David C. Parkes. 2012*a*. Peer prediction without a common prior. In *Proceedings of the 2012 ACM Conference on Electronic Commerce*. EC '12 New York, NY, USA: ACM pp. 964–981.

Witkowski, Jens and David C Parkes. 2012*b*. A robust Bayesian truth serum for small populations. In *Twenty-Sixth AAAI Conference on Artificial Intelligence*.

Witkowski, Jens and David C Parkes. 2013. Learning the prior in minimal peer prediction. In *Proceedings of the 3rd Workshop on Social Computing and User Generated Content at the ACM Conference on Electronic Commerce*. Vol. 14.

Wolfers, Justin and Eric Zitzewitz. 2006. Interpreting prediction market prices as probabilities. Technical report National Bureau of Economic Research.

The Tinbergen Institute is the Institute for Economic Research, which was founded in 1987 by the Faculties of Economics and Econometrics of the Erasmus University Rotterdam, University of Amsterdam and VU University Amsterdam. The Institute is named after the late Professor Jan Tinbergen, Dutch Nobel Prize laureate in economics in 1969. The Tinbergen Institute is located in Amsterdam and Rotterdam. The following books recently appeared in the Tinbergen Institute Research Series:

718 S. KUCINSKAS, *Essays in Financial Economics*

719 O. FURTUNA, *Fiscal Austerity and Risk Sharing in Advanced Economies*

720 E. JAKUCIONYTE, *The Macroeconomic Consequences of Carry Trade Gone Wrong and Borrower Protection*

721 M. LI, *Essays on Time Series Models with Unobserved Components and Their Applications*

722 N. CIURILĂ, *Risk Sharing Properties and Labor Supply Disincentives of Pay-As-You-Go Pension Systems*

723 N.M. BOSCH, *Empirical Studies on Tax Incentives and Labour Market Behaviour*

724 S.D. JAGAU, *Listen to the Sirens: Understanding Psychological Mechanisms with Theory and Experimental Tests*

725 S. ALBRECHT, *Empirical Studies in Labour and Migration Economics*

726 Y. ZHU, *On the Effects of CEO Compensation*

727 S. XIA, *Essays on Markets for CEOs and Financial Analysts*

728 I. SAKALAUSKAITE, *Essays on Malpractice in Finance*

729 M.M. GARDBERG, *Financial Integration and Global Imbalances*

730 U. THŰMMEL, *Of Machines and Men: Optimal Redistributive Policies under Technological Change*

731 B.J.L. KEIJSERS, *Essays in Applied Time Series Analysis*

732 G. CIMINELLI, *Essays on Macroeconomic Policies after the Crisis*

733 Z.M. LI, *Econometric Analysis of High-frequency Market Microstructure*

734 C.M. OOSTERVEEN, *Education Design Matters*

735 S.C. BARENDSE, *In and Outside the Tails: Making and Evaluating Forecasts*

736 S. SÓVÁGÓ, *Where to Go Next? Essays on the Economics of School Choice*

737 M. HENNEQUIN, *Expectations and Bubbles in Asset Market Experiments*

738 M.W. ADLER, *The Economics of Roads: Congestion, Public Transit and Accident Management*

739 R.J. DÖTTLING, *Essays in Financial Economics*

740 E.S. ZWIERS, *About Family and Fate: Childhood Circumstances and Human Capital Formation*

741 Y.M. KUTLUAY, *The Value of (Avoiding) Malaria*

742 A. BOROWSKA, *Methods for Accurate and Efficient Bayesian Analysis of Time Series*

743 B. HU, *The Amazon Business Model, the Platform Economy and Executive Compensation: Three Essays in Search Theory*

744 R.C. SPERNA WEILAND, *Essays on Macro-Financial Risks*

745 P.M. GOLEC, *Essays in Financial Economics*

746 M.N. SOUVERIJN, *Incentives at work*

747 M.H. COVENEY, *Modern Imperatives: Essays on Education and Health Policy*

748 P. VAN BRUGGEN, *On Measuring Preferences*

749 M.H.C. NIENTKER, *On the Stability of Stochastic Dynamic Systems and their use in Econometrics*

750 S. GARCIA MANDICÓ, *Social Insurance, Labor Supply and Intra-Household Spillovers*

751 Y. SUN, *Consumer Search and Quality*

752 I. KERKEMEZOS, *On the Dynamics of (Anti) Competitive Behaviour in the Airline Industry*

753 G.W. GOY, *Modern Challenges to Monetary Policy*

754 A.C. VAN VLODROP, *Essays on Modeling Time-Varying Parameters*

755 J. SUN, *Tell Me How To Vote, Understanding the Role of Media in Modern Elections*

756 J.H. THIEL, *Competition, Dynamic Pricing and Advice in Frictional Markets: Theory and Evidence from the Dutch Market for Mortgages*

757 A. NEGRIU, *On the Economics of Institutions and Technology: a Computational Approach*

758 F. GRESNIGT, *Identifying and Predicting Financial Earth Quakes using Hawkes Processes*

759 A. EMIRMAHMUTOGLU, *Misperceptions of Uncertainty and Their Applications to Prevention*

760 A. RUSU, *Essays in Public Economics*

761 M.A. COTOFAN, *Essays in Applied Microeconomics: Non-Monetary Incentives, Skill Formation, and Work Preferences*

762 B.P.J. ANDRÉE, *Theory and Application of Dynamic Spatial Time Series Models*

763 P. PELZL, *Macro Questions, Micro Data: The Effects of External Shocks on Firms*

764 D.M. KUNST, *Essays on Technological Change, Skill Premia and Development*

765 A.J. HUMMEL, *Tax Policy in Imperfect Labor Markets*

766 T. KLEIN, *Essays in Competition Economics*

767 J. YANG, *Risk and Welfare: Theory and Experiment*