Essays in Leadership and Contests

Boon Han Koh

(ORCID: 0000-0002-4893-1432)

Submitted in partial fulfilment of the requirements of the degree

Doctor of Philosophy

(Business and Economics)

December 2017

Department of Economics

Faculty of Business and Economics

The University of Melbourne
Abstract

This thesis uses laboratory experiments to explore issues in leadership and contests within economics. We make two important contributions. First, we contribute to the literature on leadership by investigating, in the first essay, the biases that members suffer from when evaluating their leader’s outcomes. We find that members attribute good outcomes of the leader more to luck and bad outcomes more to effort. Relative to the Bayesian benchmark, leaders are apportioned enough blame for their failures but given too little credit for their successes. Moreover, while the members’ beliefs depend on how the leader has been appointed, the attribution biases that we find do not tend to depend on the appointment mechanisms.

Our second contribution is towards improving our understanding of the issues pertaining to the optimal design of economic experiments within both the leadership and contest environments. Extending our design from the first study, the second essay compares two scoring rules commonly used in the literature to incentivize subjects’ belief reports. We evaluate their performance in eliciting truthful reports of the subjects’ beliefs about the decisions made by their leader. Contrary to our theoretical predictions, we find that there are no statistically significant differences in both the elicited beliefs and updating behavior between the two scoring rules.

The final essay examines the effectiveness of real-effort tasks as an experimental tool to investigate incentive effects within the contest environment. We find that subjects are motivated by non-monetary incentives in the laboratory. We provide a framework to mitigate the impact of non-monetary incentives, and find that subjects respond to monetary incentives once we mitigate the impact of these non-monetary incentives. Our framework can be useful to researchers who are interested in examining real-effort provision within a contest environment. It may also be used to study behavior in other economic environments, such as in leadership, where researchers may want to examine the factors motivating individuals to compete for leadership positions.
Declarations

This is to certify that:

1. the thesis comprises only my original work towards the Ph.D. except where indicated in the Preface;

2. due acknowledgment has been made in the text to all other material used; and

3. the thesis is fewer than 100,000 words in length, exclusive of tables, figures, bibliographies, and appendices.

Signed

Boon Han Koh
Preface

This thesis contains original research in Chapters 2 through 4.

Chapter 2 is based on the following working paper:


Chapter 3 is based on the following working paper:


Chapter 4 is based on the following publication:


All co-authorship has taken place in accordance with the Graduate Research Training Policy of the University of Melbourne.

The experiments conducted in Chapters 2 through 4 have been supported by the Australian Research Council (DP1094676). In addition, the experiments conducted in Chapter 3 have been generously supported by the Department of Economics at the University of Melbourne.
Acknowledgments

It has been an incredibly wonderful journey leading up to the submission of my Ph.D. thesis today, and if I were to list all the people that had helped me throughout in some way or another, I am sure that I would have enough material to submit a second thesis. Nonetheless, I would like to express my gratitude to the following people.

First of all, I would like to thank my advisors Nisvan Erkal, Lata Gangadharan, and Tom Wilkening, who have been extremely supportive and helpful throughout this journey. I am truly grateful to have the opportunity to learn so much from them on how to conduct solid academic research. I am especially thankful to Nisvan and Lata for being such wonderful coauthors, and for providing me with the financial support through their grants to complete these projects successfully. I am also grateful to Tom, who has always put so much faith in me ever since my Honours year.

I am deeply indebted to the Department of Economics at the University of Melbourne for their support throughout my candidature. In particular, I would like to thank Guy Mayraz who has taken so much time to provide me with feedback and comments. The Ph.D. directors have changed over the past five years at a rate that rivals the change in Australian Prime Ministers. Nonetheless, I would like to thank Christopher Skeels, Phillip McCalman, Lawrence Uren, and Kalvinder Shields for their kind words of encouragement, and for ensuring that all of the Ph.D. candidates receive the best resources they need to complete their candidature.

My colleagues at the Department have been incredible, be it to provide me with comments after each presentation, being there for moral support, or dragging me out on coffee breaks despite my countless rejections. Special shout-out to Ingrid Burfurd, Yunho (Simon) Cho, David Delacrétaz, Edwin Ip, Bryn Lampe, Todd Morris, Ellen Muir, Jia Sheen Nah, Huiwen Ng, Simeon Press, and Shannon Ward who have made this journey truly enjoyable to me.

During my candidature, I also had the wonderful opportunity to spend a few months abroad at the University of Nottingham as a Visiting Scholar. The colleagues there have been amazing in supporting me during my time at Nottingham and in providing valuable feedback for my work. They have assured me that, regardless of where we are, life as a Ph.D. candidate is never that lonely after all.
The professional staff have been incredibly helpful with assisting me on various issues pertaining to my research and travel throughout my candidature. I would like to thank, in particular, Suzie Collins and Jennifer Decolongon for supporting me on this journey, as well as Amy Corman, Michael Kerrison, Muntasha Khan, and Bing Liu for their assistance with running the experiments.

I would like to acknowledge support from the Melbourne International Research Scholarship. I am grateful to the Department of Economics and the Faculty of Business and Economics for the scholarships and travel funding I have received during my candidature.

The three core chapters of this thesis have also benefited from comments and feedback from various people that I have met, both within and outside of the University of Melbourne. Chapter 2 has greatly benefited from the comments of Kai Barron, Haley Harwell, Vai-Lam Mui, Daniele Nosenzo, Silvia Sonderegger, Georg Weizsäcker, seminar participants at WZB Berlin, the University of Nottingham, and the University of Arkansas, as well as participants at the 11th Annual Australia New Zealand Workshop on Experimental Economics (ANZWEE), the 86th Annual Conference of the Southern Economic Association (SEA), the 8th Thurgau Experimental Economics Meeting, the 2nd BEG Workshop on Behavioural and Experimental Economics (Gothenburg), the 30th PhD Conference in Economics and Business (Melbourne), and the Behavioural Economics: Foundations and Applied Research Workshop (BEFAR).

For Chapter 3, I would like to thank Maria Recalde and participants at the 12th Annual Australia New Zealand Workshop on Experimental Economics (ANZWEE) for their comments and feedback.

The publication of Chapter 4 was made possible as a result of comments from Tim Cason, Nick Feltovich, Phil Grossman, Lingbo Huang, Daniele Nosenzo, Jörg Oechssler, Roman Sheremeta, Bob Slonim, Chris Starmer, Marie Claire Villeval, the Associate Editor of the European Economic Review, three anonymous reviewers, participants at the 10th Annual Australia New Zealand Workshop on Experimental Economics (ANZWEE), the 2016 Economics and Biology of Contests Conference (EBCC), the 2016 North American Meetings of the Economic Science Association (ESA), the 2017 CBESS-CeDEx-CREED (CCC) Meeting, and seminar participants at the University of Melbourne.

My friends and family have shown me that it is also possible to enjoy life outside of academia. I am grateful to my parents, siblings, and two nephews for believing in me, and for supporting me emotionally through this arduous journey. They have definitely seen me at both my worst and best times during my candidature. I would also like to show my appreciation to my fabulous friends who have consistently provided me with the avenue I needed to de-stress on a regular basis, be in through board game sessions or simply a nice meal (and drinks).

A final shout-out to my (ex-)housemate’s cat, Dallas. Without him sitting on my papers constantly as I worked, I would probably have submitted my thesis a day earlier.

viii
## Contents

Abstract  
Declarations  
Preface  
Acknowledgments  

### 1 Introduction

### 2 Attribution biases in leadership: Is it effort or luck?

#### 2.1 Introduction

#### 2.2 Related Literature

#### 2.3 Experimental Design

- 2.3.1 Overview
- 2.3.2 Leadership task
- 2.3.3 Procedures and payment

#### 2.4 Theoretical Framework

- 2.4.1 Environment and payoffs.
- 2.4.2 Leader’s effort choice
- 2.4.3 Information and beliefs

#### 2.5 Estimation Strategy for Posterior Beliefs

#### 2.6 Results

- 2.6.1 The dictator game as a proxy for an individual’s type
- 2.6.2 Members’ interim beliefs
3 Eliciting beliefs about actions of others: A comparison between the quadratic and binarized scoring rules

3.1 Introduction

3.2 Related Literature

3.3 Experimental Design

3.3.1 Treatments

3.3.2 Procedures

3.4 Theory and Hypotheses

3.4.1 Interim beliefs

3.4.2 Updating behavior

3.5 Results

3.5.1 Decisions in the risk game

3.5.2 Interim beliefs

3.5.3 Updating behavior

3.6 Discussion

4 Monetary and non-monetary incentives in real-effort tournaments

4.1 Introduction

4.2 Related Literature

4.3 Experiment 1

4.3.1 Design and Treatments

4.3.2 Hypotheses

4.3.3 Procedures

4.3.4 Results

4.4 Experiment 2

4.4.1 Framework

4.4.2 Treatments and Hypotheses
4.4.3 Results .................................................. 100
4.5 Experiment 3 .............................................. 104
  4.5.1 Treatments and Hypotheses ......................... 104
  4.5.2 Results ............................................... 105
4.6 Conclusion .............................................. 107

5 Conclusion .............................................. 111

Bibliography .................................................. 113

A Appendices for Chapter 2 ................................. A1
  A.1 Experimental instructions .............................. A1
  A.2 Screenshots for belief elicitation task ............... A16
  A.3 Derivation of Hypothesis 2.1 ......................... A17
  A.4 Additional analysis ..................................... A21
    A.4.1 Analysis with Game 1 treatments only .......... A21
    A.4.2 Analysis of posterior beliefs with entire sample (including inconsistent and non-updaters) .......... A24
    A.4.3 IV regression of posterior beliefs ............... A25

B Appendices for Chapter 3 ................................. A27
  B.1 Experimental instructions .............................. A27
  B.2 Additional analyses of interim beliefs ............... A33

C Appendices for Chapter 4 ................................. A35
  C.1 Experimental instructions .............................. A35
  C.2 Additional analyses ................................... A60
# List of Figures

2.1 Overview of experiment .................................................. 12
2.2 Leadership task ............................................................ 15
2.3 Interpretation of $\delta$ given $Q_H$ observed and $\gamma_G = 1$ .... 25
2.4 Interpretation of $\gamma_G$ given $Q_H$ observed and $\delta = 1$ .... 27
2.5 Leader’s effort choice in Game 1 against dictator game behavior 29
2.6 Preference for leadership appointment under Treatment GA .... 31
2.7 Member’s interim belief against beliefs about leader’s behavior in the dictator game .................................................. 31
2.8 Distributions of group members’ interim beliefs ...................... 33
2.9 Interim vs. posterior beliefs .............................................. 36
2.10 Distribution of inconsistent and non-updates by subjects .......... 37
2.11 Actual vs. Bayesian posterior beliefs (less inconsistent and non-updaters) 38
3.1 Decisions in risk game (QSR vs. BSR) ............................... 63
3.2 Reported interim beliefs (QSR vs. BSR) .............................. 64
3.3 Empirical cumulative distributions of absolute distance of reported interim beliefs from 50 (QSR vs. BSR) ...................... 66
4.1 Screenshot for treatments B, BL, and BH ............................ 86
4.2 Overall effort choice by treatment (Experiment 1) ................. 92
4.3 Screenshot for treatments A1A2, A1A2-L, and A1A2-H ............ 99
4.4 Overall effort choice by treatment (Experiment 2) ............... 101
4.5 Overall effort choice by treatment (Experiment 3) ............... 106
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A.1 Investment Options (Example of a Round)</td>
<td>A6</td>
</tr>
<tr>
<td>A.2 Investment Options (Practice Question)</td>
<td>A14</td>
</tr>
<tr>
<td>A.3 Decision screens – Elicitation of beliefs</td>
<td>A16</td>
</tr>
<tr>
<td>C.1 Decision Screen for Task 1</td>
<td>A37</td>
</tr>
<tr>
<td>C.2 Encoding Activity for Task 2</td>
<td>A39</td>
</tr>
<tr>
<td>C.3 Encoding Activity for Task 2</td>
<td>A43</td>
</tr>
<tr>
<td>C.4 Confirmation Screen for Task 2</td>
<td>A45</td>
</tr>
<tr>
<td>C.5 Encoding Activity for Task 2</td>
<td>A48</td>
</tr>
<tr>
<td>C.6 Time-Out Screen for Task 2</td>
<td>A50</td>
</tr>
<tr>
<td>C.7 Encoding Activities for Task 2</td>
<td>A53</td>
</tr>
<tr>
<td>C.8 Overall effort choices by treatment (S, 20-Min, and EC)</td>
<td>A60</td>
</tr>
</tbody>
</table>
List of Tables

2.1 Order of treatments for each experiment session .................................... 19
2.2 Regression of leader’s effort choice .................................................. 30
2.3 Regression of members’ interim belief ................................................... 34
2.4 Regression of members’ posterior beliefs ............................................. 40
2.5 Finite mixture model for updating behavior .......................................... 43

3.1 Actual updating behavior (BSR) vs. implied updating behavior (QSR) .......... 61
3.2 Regression of absolute distance of interim belief from 50 ......................... 68
3.3 Regression of members’ posterior beliefs (BSR vs. QSR) ........................ 70

4.1 Summary of experimental treatments .................................................. 88
4.2 Summary statistics ............................................................................. 91
4.3 Regression results for Experiment 1 .................................................... 94
4.4 Regression results for Experiment 2 ..................................................... 103
4.5 Regression results for Experiment 3 ..................................................... 108

A.1 Regression of leader’s effort choice (Game 1) ...................................... A21
A.2 Regression of members’ interim belief (Game 1) .................................... A22
A.3 Regression of members’ posterior beliefs (Game 1) ................................. A23
A.4 Finite mixture model for updating behavior (Game 1) ............................ A24
A.5 Regression of members’ posterior beliefs (entire sample) ......................... A25
A.6 IV regression of members’ posterior beliefs .......................................... A26

B.1 Regression of members’ interim beliefs (pooled) .................................... A34
C.1 Summary of experimental treatments (including treatments S, 20-Min, and EC) ........................................... A62
C.2 Summary statistics (including treatments S, 20-Min, and EC) ................................................................. A63
C.3 Regression results for Experiment 2 (including treatments S, 20-Min, and EC) ........................................... A64
Chapter 1

Introduction

Leaders play an important role in many social, political, and economic environments. Whether a firm succeeds often depends on the managers’ ability to motivate their employees to contribute towards the organization’s goals. The policies implemented by political leaders shape the long-term economic and social outlook of their country. Hence, it is not surprising that the literature devoted to understanding the role that leadership plays in influencing the behavior of others constitutes a growing field in economics. Since the seminal paper by Hermalin (1998), economists have become increasingly interested in whether, and how, leadership can be effective in improving the welfare of the group, e.g., by overcoming free-riding or coordination problems. Some studies examine the role and ability of leaders in influencing their followers’ decisions (e.g., Potters et al., 2007; Levy et al., 2011; Gächter et al., 2012; Brandts et al., 2015), while others investigate the motivations for individuals wanting to lead despite it being costly to do so (e.g., Arbak and Villeval, 2013; Bruttel and Fischbacher, 2013; Préget et al., 2016).

This thesis examines issues in leadership by focusing on how group members evaluate leaders for their outcomes. The perceptions that members have about the leader’s outcomes can affect group interactions in several ways. For example, during an election, the electorate’s recognition of the leader’s achievements in the past year would be crucial in deciding whether the leader is reelected. In environments where leaders and their followers interact on a repeated basis, the followers’ attitude towards their leader would likely affect the leader’s ability to command authority from them in subsequent interactions.
Importantly, if the perceptions are payoff-relevant to the leaders, then understanding how leaders are judged for their outcomes is important because the evaluation of leaders can often affect the leaders’ decision-making process. For instance, if leaders’ outcomes are perceived as being driven more by luck when in fact they are driven by the leaders’ decisions, then the leaders’ incentives to exert costly effort may be lower because they may feel that they are receiving too little credit for the outcomes that they produce. Alternatively, if leaders are blamed for failures by members of their group but given too little credit for their successes, then they are likely to receive disproportionately higher penalties for failure than rewards for successes. Consequently, this may perpetuate a culture with a “fear of failure”, which may then, in turn, inhibit leaders’ incentives to take on riskier ventures that could potentially benefit the group.

Understanding the impact of and perceptions towards leadership using empirical data can be very challenging because there are often various factors that may influence the effectiveness of leaders and how they are perceived by their followers. As a result, many studies employ laboratory or field experiments to examine various issues related to leadership. One key advantage of using experiments is that it allows researchers greater control over the conditions that define the decision-making environment, thus enabling them to disentangle the various factors that affect the leader’s decisions as well as the impact of these decisions. Nonetheless, being able to execute a well-designed experiment is not a trivial task, as there are many factors that the experimenter needs to consider to ensure that the design is well-suited for the research questions they have in mind (e.g., see Bardsley et al., 2010; List et al., 2011).

This thesis makes two important contributions. First, we use controlled laboratory experiments to investigate issues in leadership. Specifically, we focus on the question of how members attribute the outcomes of their leader to their determinants, i.e., effort and luck, in an environment where the leader makes decisions under risk on behalf of the group (Chapter 2). Second, we contribute to the understanding of the tools used in economic experiments. We examine two methodological issues within the leadership and contest environments. Within leadership, we evaluate two mechanisms commonly
used by experimental economists to elicit beliefs about the decisions of others (Chapter 3). Within contests, we evaluate the use of real-effort tasks in studying incentive effects (Chapter 4). Understanding these issues is important as it can provide researchers with useful tools to advance our understanding of issues in the areas of leadership and contests. For example, Chapter 4 provides a framework that can also be valuable to researchers interested in understanding how individuals compete for leadership positions.

We begin in Chapter 2 by examining attribution biases in leadership. Leaders often have to make decisions under risk and uncertainty. Using a controlled laboratory experiment motivated by a simple theoretical framework, we investigate group members’ beliefs about the effort decisions of their leader in an environment where only outcomes are observable. We examine whether they attribute leaders’ success (or failure) to effort or to luck. We consider three different leadership appointment mechanisms: imposed, random, and group assignment. We find that members’ beliefs depend on the leadership appointment mechanism. Moreover, the members are biased when they update their beliefs upon observing the outcomes of their leaders’ decisions, but the biases do not tend to depend on the appointment mechanism. Relative to the Bayesian benchmark, group members apportion leaders the same amount of blame for their failures, but they give too little credit for their successes. In addition, members suffer from base-rate neglect in that they place too little weight on their prior beliefs as compared to a Bayesian.

Next, in Chapter 3, we use the experimental design from the previous chapter to evaluate mechanisms used in laboratory experiments to elicit subjects’ beliefs. Belief elicitation is becoming increasingly common in experimental economics. Knowing what beliefs subjects hold can help improve the experimenter’s understanding of the underlying motivations behind subjects’ decisions. Within our leadership environment, measuring the beliefs held by group members can allow us to directly examine how they evaluate the leader’s decisions. In this chapter, we evaluate the quadratic scoring rule (QSR), which is a scoring rule widely used to incentivize elicited beliefs. However, eliciting beliefs under the QSR may lead to biased conclusions if agents are not risk neutral. We investigate this issue by comparing the QSR to the binarized scoring rule (BSR) introduced by Hossain.
and Okui (2013), which is a binary lottery procedure that is incentive compatible under all assumed risk preferences. Overall, we do not find any statistically significant differences in neither the elicited beliefs nor the updating behavior between the QSR and BSR treatments.

Finally, Chapter 4 examines the effectiveness of real-effort tasks as an experimental tool to investigate contest behavior within the laboratory. Results from laboratory experiments using real-effort tasks provide mixed evidence on the relationship between monetary incentives and effort provision. To examine this issue, we design a series of experiments where subjects participate in two-player real-effort tournaments with two prizes. We find that subjects exert high effort even if there are no monetary incentives, suggesting that non-monetary incentives are contributing to their effort choices. We show that the impact of non-monetary incentives can be reduced by providing subjects with the option of leaving the laboratory early, using an incentivized timeout button, or working on an incentivized alternative activity. Using the design with an incentivized alternative activity, we show that participants increase effort in response to monetary incentives. Our findings suggest that results from real-effort tasks require a careful evaluation and interpretation of the motivations underlying the observed performance. This is particularly important if, for example, researchers would like to use real-effort tasks to understand how individuals compete for positions of authority or leadership within the laboratory.

Chapters 2 to 4 form the core of the thesis, while Chapter 5 concludes. Appendices A to C contain the supplementary material corresponding to each of the core chapters, which include the instructions used in each experiment, proofs, and additional analyses.
Chapter 2

Attribution biases in leadership: Is it effort or luck?

2.1 Introduction

Leadership is challenging because leaders often have to make decisions under risk and uncertainty. For example, politicians constantly make major policy decisions amid uncertain geopolitical climates (see, e.g., McDermott, 1998). Managers within organizations often need to decide how much of their team’s resources to devote to projects with uncertain outcomes. If effort is not observable, an important question that arises is how the performance of the leader will be evaluated. This paper studies whether there are systematic biases that characterize the evaluation process of leaders' decisions. For instance, are outcomes systematically more likely to be attributed to luck? Or, do observers put too much blame on the leader for bad outcomes? It is important to be aware of these biases because they are likely to have an impact on the decisions of the leader. If leaders are worried about being negatively judged for their failures, then they may take on less risk than optimal. For example, fear of failure is seen as one of the major impediments of innovation and corporate growth.\(^1\)

To answer these questions, we conduct a laboratory experiment motivated by a simple theoretical framework. We consider an environment where outcomes are determined by

both effort and luck, and assume that effort is not observable. Leadership is defined as the power to make decisions on behalf of a group under risk. Specifically, subjects participate in a leadership task with multiple rounds in groups of three. A leader is appointed for each group at the beginning of each round. We consider different leadership appointment mechanisms: imposed, random, and group assignment. The leadership appointment mechanism varies from round to round. The appointed leader in each round chooses between two investment options (i.e., two effort choices) with binary outcomes. The higher effort choice leads to a higher probability of the good outcome for the team but at a higher private cost to the leader. Within this design, we explore the group members’ initial beliefs about their leader conditional on the appointment mechanism and how these beliefs are updated after observing the outcome of the leader’s decision.

Standard economic theory assumes that the process of belief updating will be based on unbiased beliefs formed according to Bayes’ rule. However, studies have revealed significant deviations from Bayes’ rule across different contexts (see, e.g., Tversky and Kahneman, 1974; Grether, 1980; Eil and Rao, 2011). We consider several issues that may arise when members of the group update their beliefs about the leader. First, we study biases in the way prior beliefs are treated. That is, we ask whether members suffer from base-rate neglect (i.e., put too little weight on their prior beliefs) or confirmatory bias (i.e., put too much weight on their prior beliefs) relative to a Bayesian. Second, we examine whether, relative to a Bayesian, members respond too little or too much to new information about the action taken by the leader. Responding too little, for example, would imply that they believe luck plays a bigger role in determining outcomes. The third issue we explore is whether members treat good and bad outcomes asymmetrically. For example, if they respond more to bad outcomes than to good outcomes, this implies that they believe that a lack of effort plays a larger role in bad outcomes while luck plays a bigger role in good outcomes.

Base-rate neglect is the tendency of individuals to ignore prior information or beliefs. Kahneman and Tversky (1973) provide one of the earliest evidence of this phenomenon. On the other hand, confirmatory bias is the tendency of individuals to place too much emphasis on their prior beliefs by favoring evidence that confirms these priors and ignoring those that contradict them. See, e.g., Lord et al. (1979), Darley and Gross (1983), and Plous (1991).
Our results reveal that group members consistently suffer from base-rate neglect. Moreover, there is a statistically significant difference in the way they treat good and bad outcomes on average. The asymmetry stems from the fact that, while members attribute good outcomes more to luck as compared to a Bayesian, they respond to bad outcomes in the same way a Bayesian would. Hence, members on average attribute good (bad) outcomes more to luck (effort). As a result, leaders get too little credit for their successes.

We provide a simple theoretical framework to evaluate how members form beliefs about the leader under the different appointment mechanisms. We find that, in line with the predictions of our theoretical framework, group members take the leadership appointment mechanism into account while forming their initial beliefs about their leader’s type. As a result, they are, for example, significantly more likely to believe that the group-appointed leader will act in the group’s interest as compared to a randomly-appointed leader. However, the biases that characterize the belief-updating process do not tend to depend on the appointment mechanism.

Finally, our finite mixture model analysis uncovers heterogeneity in the members’ updating behavior. The majority of belief updates are characterized by a modest level of base-rate neglect, under-responsiveness to the leader’s outcomes, as well as an attribution of the leader’s success (failure) more to his luck (effort). On the other hand, a small proportion of updates is characterized by a much larger level of base-rate neglect as well as an over-responsiveness to the leader’s outcomes. However, this second group of members do not treat good and bad outcomes asymmetrically.

The paper proceeds as follows. We discuss our contributions to the relevant literature in the next section. We explain details of the experimental design in Section 2.3. Section 2.4 presents the theoretical framework and our hypotheses, and we provide an econometric framework that allows us to analyze the members’ updating behavior in Section 2.5. Section 2.6 discusses the results and Section 2.7 concludes.
2.2 Related Literature

Understanding biases in human judgment and behavior is important as it allows us to make the specification of the utility function more realistic than under standard economic assumptions (Rabin, 1998). As a result, for decades scholars have been interested in the different types of biases that individuals suffer from and how these biases affect behavior. Our work contributes to this literature by focusing on biases in belief formation and information processing (see, e.g., Compte and Postlewaite, 2004; Fryer et al., 2015).

There is empirical evidence to suggest that individuals may be biased in the way they treat prior information. For example, agents have been found to suffer from base-rate neglect (Kahneman and Tversky, 1973; Nisbett and Borgida, 1975) and confirmatory bias (Lord et al., 1979; Darley and Gross, 1983; Plous, 1991) across different contexts. Consistent with Kahneman and Tversky (1973) and Nisbett and Borgida (1975), we find that individuals suffer from base-rate neglect in an environment where they are asked to evaluate the outcomes of others.

Importantly, we are interested in how individuals attribute outcomes of others to effort and luck, and whether they attribute these outcomes asymmetrically. This question has roots within the psychology literature on attribution theory (see, e.g., Miller and Ross, 1975). More recently, studies in experimental economics have also focused on this issue, but most of the attention has been on ego-related beliefs (Eil and Rao, 2011; Ertac, 2011; Grossman and Owens, 2012; Möbius et al., 2014). Both Eil and Rao (2011) and Möbius et al. (2014) find evidence of asymmetric updating where agents are more responsive to good news than to bad news about their own performance in an IQ test or a beauty task. Ertac (2011) observes the opposite effect while Grossman and Owens (2012) find no such effect. Our work differs from these studies because we do not focus on beliefs about

---

3E.g., see Tversky and Kahneman (1974) and Grether (1980). More recent studies have devoted their attention to understanding how biases in beliefs can explain why agents exhibit overconfidence, optimism/pessimism, or wishful thinking (see, e.g., Bénabou and Tirole, 2002; Brunnermeier and Parker, 2005; Köszegi, 2006; Bracha and Brown, 2012; Mayraz, 2014).

4Ajzen (1977) finds that individuals do not suffer from such a bias. Tversky and Kahneman (1982) argue that whether the base-rate information is used or neglected largely depends on the context. Stanovich and West (2008) finds that base-rate neglect is uncorrelated with cognitive ability.

5For a theoretical model of confirmatory bias, see Rabin and Schrag (1999).
an individual’s own ability or physical attributes where one’s ego can play a big role in shaping these beliefs. We contribute to this literature by considering the evaluation of the performance of other individuals in an environment where the individuals’ performance is payoff-relevant to the observers. We find strong evidence of a ‘bad news’ effect in belief updating, in that the group members respond more to bad outcomes than good outcomes of their leader. This result is in contrast to Eil and Rao (2011) and Möbius et al. (2014), but it is in line with Grossman and Owens (2012).

Two recent papers most related to our work are Barron (2016) and Coutts (2017), in that they also do not focus (solely) on ego-related beliefs. Barron (2016) considers an environment where individuals update their beliefs about the composition of the urn, i.e., where states are only financially-relevant to the agents. Coutts (2017) focuses on belief updating across a broad set of contexts (ego-related beliefs as well as beliefs about both financially-relevant and value-neutral states of the world). However, neither of these studies consider an environment where the realized states are determined by the decisions made by others, which is what we focus on in this paper. Nonetheless, Coutts (2017) provides complementary findings to our paper in that individuals tend to respond more to bad news than to good news. On the other hand, like us, Barron (2016) finds heterogeneity in updating behavior, although none of the sub-groups identified in his sample suffer from asymmetry in updating.

Our work is also related to the literature examining, using a principal-agent framework, how individuals assign blame and/or credit for outcomes arising from the decisions made by others under risk or uncertainty. There is evidence from laboratory experiments that individuals suffer from outcome bias, in that principals are biased towards one outcome over the other despite being able to fully observe the agents’ decisions. In particular, principals tend to respond to the agents’ bad outcomes or luck either by punishing them or by rewarding them less as compared to when good outcomes are realized (e.g., Gurdal et al., 2013; König-Kersting et al., 2016; Brownback and Kuhn, 2017; de Oliveira et al., 2017).6 Our findings can also be applied to this environment where the leaders in our

6Outside of the laboratory, there is also empirical evidence to suggest that such an outcome bias exists, e.g., when it comes to CEO compensation (Bertrand and Mullainathan, 2001; Leone et al., 2006) or in
set-up are the agents and the members are the principals. However, unlike these studies, we focus on the evaluation of outcomes in an environment where the leader’s decision is unobservable. Moreover, we strip away any strategic incentives for the leaders to influence the members’ decisions through their own actions, and focus primarily on how members evaluate their leaders based on the outcomes that they observe. In particular, we focus on biases that members suffer from relative to the theoretical Bayesian benchmark.

Finally, our work also contributes to the literature on leadership, a topic that has gained much attention within economics in recent years. The main focus in this growing literature has been the role and ability of the leader in affecting the decisions of others (see, e.g., Brandts et al., 2015; Gächter et al., 2012; Hamman et al., 2011; Kulas et al., 2013; Levy et al., 2011). We extend and contribute to this literature with our focus on the group members’ attribution biases. Hence, unlike the existing studies that focus on the leader’s behavior and how they can influence the behavior of their followers, we are more interested in how members of the group perceive the decisions of the leaders. One exception is the recent study by Grossman et al. (2016), who examine how followers evaluate the effectiveness of leaders and compensate them for their outcomes. However, their focus is on the differences between perceptions towards male and female leaders. They find that female leaders are less likely to be credited for group success.

2.3 Experimental Design

2.3.1 Overview

Figure 2.1 presents an overview of the experiment. The main task in the experiment is the leadership task, which we explain in detail in Section 2.3.2. The experiment features a within-subject treatment design, where subjects play six repeated rounds of the leadership task. In each round, subjects are re-matched to a new group of three. We adopt a perfect sports (Gauriot and Page, 2017). Also related to this discussion is the study by Cappelen et al. (2013) who find that individuals favor redistribution more if inequalities are due to luck but much less if the ex-post inequalities are due to decisions made under risk by the individuals themselves. Another related study is Rey-Biel et al. (2017), who examine how redistribution decisions depend on the determinants of income from a real-effort task, but their focus is on the differences in behavior between Spanish and American subjects.
stranger matching protocol, i.e., each subject is never matched with the same person more than once during the experiment. In each group, there is a leader who makes an investment decision on behalf of the group.\(^7\) The other group members do not observe the leader’s decision, and only the leader’s outcomes are observable. The key variables of interest are the members’ beliefs about their leader’s investment decisions.

All of the decisions are elicited using the strategy method, and subjects are informed whether they had been assigned the role of the leader at the end of the experiment. Hence, in each round, all subjects make their investment decisions assuming that they have been assigned to be the leader, and then state their beliefs about their leader’s investment decision assuming that someone else in the group has been assigned to be the leader.\(^8\) No feedback is given at any time during the entire experiment.

At the beginning of the experiment and prior to the leadership task, subjects first play the dictator game in groups of two. Each subject is given 300 Experimental Currency Units (ECU) and is asked to allocate this endowment between themselves and their matched partner. Subjects also make this decision using the strategy method, i.e., both subjects within the pair make their allocation decisions as the dictator. They are told that one of the decisions will be randomly chosen at the end of the experiment to determine the final allocation of the given endowment within each pair. Once subjects have played the dictator game, they receive instructions for the leadership task.\(^9\)

\(^7\)We specifically choose to use a stated-effort mechanism (as opposed to using a real-effort task) to elicit the leader’s effort choice. Under this mechanism, the outcome of the leader only has two determinants: (i) the leader’s effort choice, which is entirely within the leader’s control; and (ii) luck, which is entirely outside of the leader’s control (but can be observed by the experimenter). Using a real-effort task would have added an additional determinant, i.e., an individual’s natural ability to excel in the task, which is an unobservable characteristic to the experimenter. Given that we are interested in understanding the extent to which members believe leaders to be responsible for their outcomes, we use a stated-effort mechanism so that we can fully observe and control for the determinants of outcomes in our econometric analyses.

\(^8\)One concern with the use of the strategy method is that members’ responses may be considered to be too “cold” for them to be a realistic representation of the natural environment (as compared to the direct-response method where members make “hot” responses after observing the actual outcomes of the leader). Brandts and Charness (2011) examine this issue and find that, among the studies that they surveyed, there are more studies that find no difference in experimental results between the two methods than studies that find a difference. They also find that differences between the two elicitation methods are more likely to occur when there are fewer contingent choices to be made under the strategy method. In our experiment, subjects make a total of 31 contingent choices in the entire experiment. Given this, it is unlikely that the use of the strategy method will significantly affect the type of responses that we observe.

\(^9\)The instructions can be found in Appendix A.1.
Figure 2.1: Overview of experiment

- Dictator game in groups of two.
- Receive instructions for leadership task.
- Practice round for leadership task.
- Impose common prior by revealing average dictator game behavior from other studies in the literature.
- Indicate preferences for leader’s type for treatment GA.

**Round 1 of Leadership Task**

- **Stage 1**
  Leadership appointment mechanism revealed.
  State beliefs about other group members’ preferences for leader (Treatment GA only).

- **Stage 2**
  Make investment decisions as leaders.
  State beliefs about appointed leader’s decision in dictator game.

- **Stage 3**
  State interim and posterior beliefs about the leader’s effort decision.

**Rounds 2 to 6 of Leadership Task**

- Post-experimental questionnaire.
- Realization of outcomes and payments.
2.3.2 Leadership task

Each round of the leadership task consists of three stages. In Stage 1, we reveal the leadership appointment mechanism to the subjects. In Stage 2, all subjects make their investment decisions assuming that they have been appointed to be the leader. Finally, in stage 3, all subjects are asked to state their beliefs about the leader’s effort decisions, assuming that they have not been appointed to be the leader.

Leadership appointment. In the first stage, subjects are informed that one of the group members is appointed to be the leader. There are three possible appointment mechanisms across four treatments, and subjects are informed how the leader will be appointed for that round (although they are not told who is appointed to be the leader). The leadership appointment mechanism varies across the six rounds. However, in each round, the appointment mechanism is the same for all the subjects within the same session.

Our choice of the different appointment mechanisms described below is motivated by the fact that, in reality, leaders are appointed in a variety of ways. We consider different appointment mechanisms to examine whether members’ beliefs about the leader’s actions prior to observing their outcomes depends on the way the leader has been appointed. Moreover, if the appointment mechanisms have an impact on these prior beliefs, then we can also examine how variations in these prior beliefs affect the members’ updating behavior. If it is the case that biases depend on the appointment mechanism (e.g., members are more likely to blame leaders for their failures if the leader was not appointed by the group), then an important question is whether there is an optimal mechanism that minimizes the biases that members hold in their beliefs about their leader’s outcomes.

The leadership appointment mechanisms in the first three treatments are exogenous in that they are not within the direct control of the members in each group. Under the random assignment mechanism (treatment RA), an individual is randomly selected to be the leader and each individual has an equal chance of being appointed. This treatment serves as the baseline for all treatment comparisons in our experiment. In two separate
treatments, leadership is imposed on the group. Specifically, in treatments LA and HA, the group member who contributed the least and highest amount to their matched partner in the dictator game, respectively, is appointed to be the leader.\textsuperscript{10}

The fourth mechanism is the group appointment mechanism (treatment GA), where leadership appointment is determined by the members of the group. Before beginning the first round of the leadership task, each group member is asked to indicate whether they prefer: (i) to appoint the member who contributed the least to their matched partner in the dictator game; (ii) to appoint the member who contributed the most to their matched partner in the dictator game; or (iii) to randomly select one member to be the leader. Subjects make this decision once, and they are informed that their indicated preference will be used whenever the leader is appointed under treatment GA in any given round.

If treatment GA is implemented in a given round, then one of the decisions of the group members is randomly picked by the computer and one of the \textit{other} two group members is appointed to leadership based on this randomly selected decision. This ensures that there is no scope for strategic behavior in the nomination decisions in that subjects are unable to influence their own probability of being the leader through their own nomination decision.\textsuperscript{11} This is especially important in our set-up because there is a clear advantage from being the leader. As can be seen later, the appointed leader will always receive a strictly higher payoff than each of the other group members.

Given that each nomination decision within the group is equally likely to be used to determine the leadership appointment in treatment GA, the beliefs that subjects hold about their group members’ nomination decisions may also play a role in shaping their beliefs about their leader’s decision in this treatment. Hence, in the round where treatment GA is implemented, we also ask subjects to state what they believe the other two group members indicated as their preference for the appointed leader. They are paid an additional 10 ECU if their guess is correct.

\textsuperscript{10}The details of the leadership task are only revealed after the subjects have made their decisions in the dictator game. This ensures that any strategic behavior in the dictator game is minimized. It is important to note that the actual decisions of each subject in the dictator game are never revealed to their group members in the leadership task at any point during the experiment.

\textsuperscript{11}See, e.g., Galeotti and Zizzo (2015), for a similar protocol.
Leader’s investment decision. In the second stage, each subject is asked to make an investment decision on behalf of the group, assuming that s/he has been appointed to be the leader. The subject’s decision in a given round is only implemented if s/he is appointed to be the leader of the group in that round.

Subjects are informed that the appointed leader is given an individual endowment of 300 ECU, and that the leader will have to choose between two investment options that will affect the payoffs of all the group members. The two investment options, given in Figure 2.2, are: (i) Investment X, which corresponds to a high effort level; and (ii) Investment Y, which corresponds to a low effort level. Both investment options yield the same high return if they succeed and the same low return if they fail. However, they differ both in their probability of success and failure, and in their cost to the leader. Investment X succeeds with a probability of 0.75 and costs the leader 250 ECU, while Investment Y succeeds with a probability of 0.25 but costs the leader only 50 ECU.

![Figure 2.2: Leadership task](image)

It is possible that the subjects’ investment decisions (as the leader) and their beliefs (as members) about the investment decision of the leader are sensitive to the payoffs
associated with each investment option. For instance, some subjects may be averse to choosing Investment Y as the leader if the investment options are such that their members will receive zero payoffs when the investment fails.\textsuperscript{12} Hence, in each round, subjects are faced with either Game 0 or Game 1, where the payoffs associated with both investment options vary between the two games. For Game 0, the investment provides a return of 600 ECU if it succeeds and 0 ECU if it fails. For Game 1, these payoffs are 750 ECU and 150 ECU, respectively. The return from the investment is distributed evenly between the leader and the two members. The leader’s final payoff (given at the bottom of Figure 2.2) is the net of her given endowment, less the cost of investment, plus her share of the return from the investment. On the other hand, each member receives only his share of the return from the investment as the final payoff.

Finally, subjects are informed that the leader’s investment decision will not be revealed to their group members. Instead, the group members will only be informed whether the investment has succeeded or failed at the end of the experiment.

**Elicitation of beliefs of group members.** In the third and final stage, subjects are asked to state their beliefs of the likelihood that their leader has chosen Investment X (i.e., high effort), assuming that they have not been assigned to be the leader. We elicit two sets of beliefs from each subject. First, each subject is asked to state their unconditional belief that the leader has chosen high effort. Given that the subjects form these beliefs after being informed of the leadership appointment mechanism, we refer to these unconditional beliefs as the members’ *interim* beliefs. Second, each subject is asked to state their belief that the leader has chosen Investment X, conditional on observing whether the investment has succeeded or failed. We refer to these beliefs as the members’ *posterior* beliefs.

We elicit both sets of beliefs on two separate screens.\textsuperscript{13} On the first screen, each subject is asked what they think is the likelihood that their leader has chosen Investment X. However, some subjects may not be fully comfortable with the concept of probabilities

\[\text{\textsuperscript{12}For example, Incekara-Hafalir and Stecher (2016) find that some individuals violate expected utility theory in Allais-type tasks due to an aversion to receiving a zero outcome.}\]

\[\text{\textsuperscript{13}Screenshots of the decision screens can be found in Appendix A.2.}\]
and chance. Moreover, subjects have been found to perform better in terms of additivity and Bayesian updating when beliefs are elicited as a population frequency (Schlag et al., 2015). Hence, in our experiment, we ask subjects to consider a hypothetical situation where there are 100 people in the position that their leader is currently in. We then ask them to indicate how many of these individuals they think would choose Investment X. When stating their beliefs, subjects are required to enter an integer number between 0 and 100.

On the second screen, subjects are asked to state what they think is the likelihood that their leader has chosen Investment X, but under two different scenarios: (i) Suppose that the investment chosen by their leader has succeeded, which provides us with the members’ posterior beliefs conditional on observing a good outcome; and (ii) Suppose that the investment chosen by their leader has failed, which provides us with the members’ posterior beliefs conditional on observing a bad outcome. We elicit the posterior beliefs in exactly the same way as the interim beliefs. That is, we ask subjects to consider a hypothetical situation where there are 100 people in the position that their leader is currently in. We then ask them to indicate how many of these individuals they think would choose Investment X, conditional on observing the outcome of the leader’s investment decision. We provide the subjects with the interim belief that they have stated in the previous screen and ask them to consider whether their posterior beliefs are the same as, or different to, their interim belief under each scenario. However, we do not impose any restrictions on their posterior beliefs. In other words, the group members can state any belief they want, regardless of what their interim beliefs are.

Beliefs are incentivized using the Binarized Scoring Rule (BSR). The BSR modifies a proper scoring rule (such as the Quadratic Scoring Rule, QSR) using a binary lottery procedure such that the subject receives a fixed reward with a probability that is determined by the relevant loss function. Specifically, under this scoring rule, the distance between the subject’s belief report and the actual investment decision of the leader determines his probability of receiving 10 ECU. The further the subject’s reported belief is from the
leader’s investment decision, the lower the probability of receiving this fixed payment.\footnote{Specifically, for a given belief report \( r \in [0, 100] \), the group member receives 10 ECU with probability 
\[ 1 - \left[ I(e = e_H) - \frac{r}{100} \right]^2, \]
where \( I(e = e_H) \) is an indicator variable that equals 1 if the leader chose high effort (Investment X) and 0 otherwise.} We adopt the BSR because it incentives truth-telling independent of the subjects’ risk preferences (Hossain and Okui, 2013). Subjects are paid for either their interim belief or their posterior beliefs. Since posterior beliefs are elicited using the strategy method, if subjects are paid for their posterior beliefs, then they will be paid for the stated belief that corresponds to the actual outcome of the investment chosen by their leader.\footnote{Hence, if the investment chosen by the leader has succeeded, then the members are paid for their reported posterior belief conditional on good outcome, and vice versa.}

In addition to the belief questions stated above, prior to the belief-elicitation stage, subjects are also asked what they think their appointed leader transferred to their matched partner in the dictator game. The options given are: (i) 0 ECU; (ii) 1-50 ECU; (iii) 51-150 ECU; (iv) 151-200 ECU; (v) 201-250 ECU; and (vi) 251-300 ECU. They are asked to answer this question under the assumption that they have not been appointed to be the leader, and are paid an additional 10 ECU if their guess of the leader’s decision is correct.

\subsection*{2.3.3 Procedures and payment}

The experiments were conducted in the Experimental Economics Laboratory at the University of Melbourne (\( E^2MU \)) and programmed using z-Tree (Fischbacher, 2007). Subjects were recruited using ORSEE (Greiner, 2015), and we restricted our sample to Australian citizens.\footnote{Cultural differences may also affect individuals’ beliefs about the determinants of outcomes (e.g., see Alesina and Angeletos, 2005; Rey-Biel et al., 2017). As such, we restricted our sample to Australian citizens only to reduce the impact that differences in culture may have on the members’ belief-updating process.} We ran a total of 10 sessions with 24 to 30 subjects in each session. A total of 282 subjects participated in the experiment.\footnote{To ensure consistency of our subject pool, we decided to drop 10 subjects from the analysis. Two subjects had prior experience with the experiment, while eight subjects had misreported their citizenship on the recruitment system and indicated in the questionnaire that they have lived in Australia for less than two years. Hence, data from 272 subjects are used for the analysis.}

Table 2.1 summarizes the order of the treatments implemented in each session. In each cell of the table, the first two letters denote the leadership appointment mechanism,
while the Arabic numeral at the end denotes the game implemented for the corresponding round in the session. For example, a cell that states “RA1” means that the leader was appointed randomly for that round (treatment RA), and the subjects played Game 1.

To ensure that the subjects fully understood the tasks, the experimenter verbally summarized the instructions after they had finished reading the printed instructions. Before beginning the actual leadership task, subjects completed a set of control questions and participated in a practice round using treatment GA and Game 0. To reduce experimental fatigue, we implemented only six rounds of the leadership task. We implemented all four appointment mechanisms for Game 1 and treatments LA and HA for Game 0. This allows us to study the subjects’ behavior across different mechanisms using the same set of parameters. We chose to implement treatments LA and HA for Game 0 as the theoretical difference in interim beliefs between these two treatments is the greatest (explained in detail in the next section).

The orders between treatments LA and HA, as well as that between treatments RA and GA, were changed to control for potential order effects. Note that both the appointment mechanism and the payoffs from the investment options could vary from round to round during the experiment. Hence, in order to avoid confusing subjects with changes in the task on multiple dimensions early on in the experiment, we always implemented Game 0 in Rounds 1 and 2 (as in the practice round) before implementing Game 1 in Rounds 3 to 6.\(^{18}\)

Each session lasted between 90 and 120 minutes. At the end of the experiment, sub-

\(^{18}\)Note that, as a result of this design choice, any conclusions regarding the subjects’ behavior in Game 1 should be interpreted as conditional on their experience with Game 0.
jects were invited to complete a brief questionnaire which included demographic questions, questions about their decisions during the experiment, and an incentivized one-shot risk game (Gneezy and Potters, 1997) to elicit their risk preferences. Subjects were paid for either the dictator game or the leadership task. If they were paid for the leadership task, then we paid them for their decisions in one of the six rounds. For the chosen round, a leader was appointed according to the corresponding treatment and the leader was paid only for their investment decision. The other two members were paid for their leader’s decision as well as their stated beliefs. Earnings were converted to cash at the conclusion of the session at the rate 10 ECU = 1 AUD. Overall, subjects earned between $10 and $76, with the mean earnings being $34.07. Subjects’ earnings also included a show-up fee of $10.

2.4 Theoretical Framework

We provide a simple theoretical framework to evaluate how the group members form beliefs about their leader given the leadership appointment mechanism, and subsequently, how they update their beliefs upon observing the leader’s outcome.

2.4.1 Environment and payoffs.

Players maximize expected utility and are differentiated based on their other-regarding preferences. Let $\beta_i \geq 0$ denote the type of player $i$ and it is a private draw from a distribution $F(\beta)$ with density $f(\beta)$. $F(\beta)$ is common knowledge. Players are randomly assigned to groups of size $N$. One player in each group is assigned to be the leader of the group. The leader makes a decision that determines the payoffs of everyone in the group. Specifically, the leader makes an effort choice $e \in \{e_L, e_H\}$ at cost $c \in \{c_L, c_H\}$, which is deducted from an initial endowment $\omega$ that the leader receives. Assume that $\omega \geq c_H > c_L > 0$. There are two possible team outputs, $Q \in \{Q_L, Q_H\}$, where $Q_H > Q_L$, and the leader’s effort choice determines the probability with which each output level will be realized. A high effort choice leads to the high output level with a higher probability,
but it costs more to the leader. Specifically, a high effort choice \( e_H \) leads to an output \( Q_H \) with probability \( p \) and \( Q_L \) with probability \( 1 - p \), where \( p \in (0.5, 1) \). Conversely, a low effort choice \( e_L \) leads to an output \( Q_L \) with probability \( p \) and \( Q_H \) with probability \( 1 - p \).

We assume that the realized outcome is equally shared between the group members although the cost of effort is a private cost solely borne by the leader. Hence, each member in the group receives \( \frac{Q}{N} \). For any realized outcome \( Q \in \{Q_L, Q_H\} \), the utility of the leader is given by

\[
U = u \left( \frac{Q}{N} + \omega - c \right) + \beta \cdot v \left( \frac{(N - 1)Q}{N} \right).
\]

(2.1)

\( u(\cdot) \) and \( v(\cdot) \) are twice differentiable utility functions, where \( u'(\cdot), v'(\cdot) > 0 \) and \( u''(\cdot), v''(\cdot) \leq 0 \). \( u(\cdot) \) represents the direct utility of the leader from her own monetary payoff while \( v(\cdot) \) is the utility derived by the leader from the total payoffs of the members.\(^{19}\) \( \beta \) represents the type of the leader and determines the weight the leader puts on the utility of the other group members.\(^{20}\)

### 2.4.2 Leader’s effort choice

The leader maximizes her expected utility and chooses \( e_H \) over \( e_L \) if \( EU(e_H) \geq EU(e_L) \), i.e.,

\[
p \left[ u \left( \frac{Q_H}{N} + \omega - c_H \right) + \beta v \left( \frac{(N - 1)Q_H}{N} \right) \right] + (1 - p) \left[ u \left( \frac{Q_L}{N} + \omega - c_H \right) + \beta v \left( \frac{(N - 1)Q_L}{N} \right) \right]
\geq (1 - p) \left[ u \left( \frac{Q_H}{N} + \omega - c_L \right) + \beta v \left( \frac{(N - 1)Q_H}{N} \right) \right] + p \left[ u \left( \frac{Q_L}{N} + \omega - c_L \right) + \beta v \left( \frac{(N - 1)Q_L}{N} \right) \right].
\]

(2.2)

\(^{19}\)Note that this specification assumes leaders only care about the payoff that members receive from the realized outcome \( Q \). This payoff is directly within the leader’s control through her effort choice. During the experiment, members also receive a payment for their reported beliefs. However, the leader does not have direct control over the belief payment since members are paid for the accuracy of their beliefs (although it is possible that s/he has an indirect influence over the members’ beliefs through her own actions). Moreover, this payment is small relative to the payoff that members receive from the realized outcome \( Q \). Consequently, we assume that a leader with \( \beta > 0 \) is more concerned with the members’ payoff that is a direct result of her own actions. Hence, the members’ belief payments during the experiment is not included in (2.1).

\(^{20}\)See Rotemberg (2014) for a survey of models of social preferences used in the literature.
In our experiment, the choice of parameters for both Game 0 and Game 1 correspond to $N = 3$, $\omega = 300$, $p = 0.75$, $Q_H = 750$ (Game 1) / 600 (Game 0), $Q_L = 150$ (Game 1) / 0 (Game 0), $c_H = 250$, and $c_L = 50$. This implies that $\frac{Q_H}{N} + \omega - c_L > \frac{Q_L}{N} + \omega - c_H$. If $\beta = 0$, the leader only cares about her own payoffs and chooses $e_L$ since $EU(e_H) - EU(e_L) = p \left[ u \left( \frac{Q_H}{N} + \omega - c_H \right) - u \left( \frac{Q_L}{N} + \omega - c_L \right) \right] + (1 - p) \left[ u \left( \frac{Q_L}{N} + \omega - c_H \right) - u \left( \frac{Q_H}{N} + \omega - c_L \right) \right] < 0$.

For $\beta > 0$, (2.2) holds if

$$\beta \geq \frac{p \left[ u \left( \frac{Q_H}{N} + \omega - c_H \right) - u \left( \frac{Q_L}{N} + \omega - c_L \right) \right] + (1 - p) \left[ u \left( \frac{Q_L}{N} + \omega - c_H \right) - u \left( \frac{Q_H}{N} + \omega - c_L \right) \right]}{(2p - 1) \left[ v \left( \frac{(N-1)Q_H}{N} \right) - v \left( \frac{(N-1)Q_L}{N} \right) \right]} \equiv \beta^*.$$  

(2.3)

Intuitively, the leader chooses high effort if she cares sufficiently about the payoffs of the other group members. In the experiment, subjects’ decisions in the dictator game provide a proxy for their types ($\beta_i$). To impose a common prior about the distribution of types, we use Engel (2011)’s survey of standard dictator games. As indicated in Figure 2.1, subjects were informed of the average behavior observed in the literature at the beginning of the leadership task.

### 2.4.3 Information and beliefs

Based on the leadership appointment, each member forms an *interim* belief about his leader’s type. The members then update their interim beliefs about their leader’s type after observing the outcome. These updated beliefs are known as the members’ *posterior* beliefs.

**Members’ interim beliefs.** We first consider the members’ interim beliefs about their leader’s type after observing the leadership appointment mechanism. Specifically, we are interested in the member’s belief that the leader is of type $\beta \geq \beta^*$, which corresponds to the likelihood that the leader chooses $e_H$ over $e_L$. We denote member $i$’s *interim* belief upon observing leadership appointment mechanism $\Psi$ as $\mu^\Psi_i$.

Our first testable prediction is about the ranking of the members’ interim beliefs
under the different leadership appointment mechanisms. We first show in Appendix A.3 that regardless of their type, all players will prefer to have the highest type appointed as the leader. Intuitively, this is because all group members will want the leader to choose the higher effort level $e_H$, which maximizes the expected payoffs of the members. We then show in Appendix A.3 that, for any general distribution of $\beta$, this leads us to the following hypothesis:

**Hypothesis 2.1:** For any given member with type $\beta_i$, $\mu_i^{LA} < \mu_i^{RA} < \mu_i^{HA}$ and $\mu_i^{RA} < \mu_i^{GA}$.

**Members’ posterior beliefs.** We next consider how the members update their beliefs about their leader’s type after observing the outcome. The outcome $Q \in \{Q_L, Q_H\}$ is a signal that members receive that is correlated with the leader’s type. Note that $Pr(Q_L|\beta < \beta^*) = Pr(Q_H|\beta \geq \beta^*) = p$ and $Pr(Q_L|\beta \geq \beta^*) = Pr(Q_H|\beta < \beta^*) = 1 - p$. Hence, we can think of $p$ as the objective probability of receiving a signal that is correlated with the leader having type $\beta \geq \beta^*$.

We first consider the case where the group member is fully Bayesian when updating his belief conditional on observing the leader’s signal. Denote the *unbiased posterior belief* of a Bayesian group member $i$, given a signal $Q$, as $\phi_i^\Psi|Q$. In other words, $\phi_i^\Psi|Q$ is the member’s posterior belief that the leader is of type $\beta \geq \beta^*$ conditional on observing a signal $Q \in \{Q_L, Q_H\}$ as derived using Bayes’ rule.

Suppose the members receive a signal $Q = Q_H$. Using Bayes’ rule, member $i$’s posterior belief about the leader’s type is given by

$$\phi_i^\Psi|Q_H = \frac{\mu_i^\Psi \cdot Pr(Q_H|\beta \geq \beta^*)}{Pr(Q_H)} = \frac{\mu_i^\Psi p}{\mu_i^\Psi p + (1 - \mu_i^\Psi)(1 - p)}.$$

We can express the Bayesian group member’s posterior beliefs in terms of a log likelihood ratio, conditional on observing the signal $Q = Q_H$. Hence, we have

$$\log\left(\frac{\phi_i^\Psi|Q_H}{1 - \phi_i^\Psi|Q_H}\right) = \log\left(\frac{\mu_i^\Psi}{1 - \mu_i^\Psi}\right) + \log\left(\frac{p}{1 - p}\right). \quad (2.4)$$
Similarly,
\[
\log\left(\frac{\phi_i^\Psi|Q_L}{1 - \phi_i^\Psi|Q_L}\right) = \log\left(\frac{\mu_i^\Psi}{1 - \mu_i^\Psi}\right) + \log\left(\frac{1 - p}{p}\right). 
\] (2.5)

By letting \( \text{logit}(x) \equiv \log\left(\frac{x}{1-x}\right) \), we can jointly express (2.4) and (2.5) as
\[
\text{logit}(\phi_i^\Psi|Q) = \text{logit}(\mu_i^\Psi) + I(Q = Q_H) \cdot \logit(p) + I(Q = Q_L) \cdot \logit(1 - p), 
\] (2.6)

where \( I(\cdot) \) is an indicator function.

We test the null hypothesis that the members will be unbiased (i.e., Bayesian) when they update their beliefs. This allows us to determine any kind of biases they may have in their belief updating process. Moreover, our design allows us to examine whether the members’ updating behavior depends on the leadership appointment mechanism. In other words, we can observe, for example, whether being appointed by the group (treatment GA) has an impact on the way the members update their beliefs about the leader. We summarize our hypothesis as follows.

**Hypothesis 2.2:**

(i) Group members behave like Bayesian agents when updating their beliefs about the leader.

(ii) Group members behave like Bayesian agents under all of the appointment mechanisms.

In the next section, we provide an econometric framework that allows us to test Hypothesis 2.2 empirically.

### 2.5 Estimation Strategy for Posterior Beliefs

We now consider a member who is biased in his updating of beliefs about the leader’s type upon observing the signal \( Q \). Following Möbius et al. (2014), we augment the belief updating equation in (2.6) in the following way:
\[
\text{logit}(\hat{\phi}_i^\Psi|Q) = \delta \text{logit}(\mu_i^\Psi) + \gamma_G I(Q = Q_H) \cdot \logit(p) + \gamma_B I(Q = Q_L) \cdot \logit(1 - p), 
\] (2.7)
where $\hat{\phi}|_Q$ is the *biased posterior belief* of a non-Bayesian member $i$, conditional on observing $Q \in \{Q_L, Q_H\}$. The parameters $\delta$, $\gamma_G$, and $\gamma_B$ in (2.7) allow us to capture non-Bayesian updating by the biased group member. Note that $\delta = \gamma_G = \gamma_B = 1$ equates (2.7) to (2.6). This is the case where the member is fully Bayesian and does not suffer from any bias in belief updating. Hence, an evaluation of Hypothesis 2.2 is to test the null hypothesis that $\delta = \gamma_G = \gamma_B = 1$, both (i) at the pooled level; and (ii) for each appointment mechanism.

We now consider various channels through which deviations from the Bayesian benchmark can occur. $\delta$ captures the weight that the biased group member places on his interim belief when updating his posterior, $\gamma_G$ captures the extent to which the member responds to a signal of good outcome from the leader, and $\gamma_B$ captures the extent to which the member responds to a signal of bad outcome from the leader. We can think of $\gamma_G$ and $\gamma_B$ as telling us the extent to which a group member attributes a given outcome from the leader to her effort choice, relative to a Bayesian.

![Figure 2.3: Interpretation of $\delta$ given $Q_H$ observed and $\gamma_G = 1$](image)

We first focus on the weight placed on the interim beliefs, i.e., $\delta$. Figure 2.3 shows the...
implications of different values of $\delta$ on the relationship between the member’s posterior and interim beliefs, conditional on observing a good outcome and holding $\gamma_G$ constant (at 1). Note that $\delta$ governs the slope of the linear regression.

If $\delta < 1$, then the member suffers from base-rate neglect in that he ignores his interim belief when faced with signals that contradict with this prior information. To see this, consider a member with $\delta < 1$ whose interim belief $\mu_A$ is less than 0.5. This corresponds to $\text{logit}(\mu_A) < 0$ in Figure 2.3. Hence, the member believes that his leader is more likely to have chosen low effort. When $Q_H$ is observed, the signal that the member receives contradicts with his interim belief. However, he arrives at a posterior belief that is greater than that of a Bayesian (i.e., point $A'$ instead of point $A$). In other words, the member neglects his interim belief and over-updates his belief in response to receiving a signal that contradicts with what he initially believes to be true. On the other hand, now consider a member with $\delta < 1$ whose interim belief $\mu_B$ is greater than 0.5, which corresponds to $\text{logit}(\mu_B) > 0$ in Figure 2.3. Hence, this member believes that his leader is more likely to have chosen high effort. When $Q_H$ is observed, the signal that the member receives confirms his interim belief. Yet, he arrives at a posterior belief that is lower than that of a Bayesian (i.e., point $B'$ instead of point $B$). This implies that the member under-updates his belief when he receives a signal that confirms what he previously thinks to be true. One can also think of this as the member neglecting his interim belief even though he receives a signal that confirms this belief.

Conversely, $\delta > 1$ implies that the member suffers from confirmatory bias, in that he places too much weight on signals that confirms his interim beliefs. To see this, now consider a member with $\delta > 1$ whose interim belief $\mu_B$ is greater than 0.5, i.e., $\text{logit}(\mu_B) > 0$ in Figure 2.3. When $Q_H$ is observed, the signal that the member receives confirms his interim belief. He then arrives at a posterior belief that is greater than that of a Bayesian (i.e., point $B''$ instead of point $B$). In other words, the member over-updates his belief relative to a Bayesian when he receives a signal that confirms what he initially believes to be true. On the other hand, now consider a member with $\delta > 1$ whose interim

---

21 A similar analysis can be done for the case where a bad outcome is observed.
belief $\mu_A$ is less than 0.5, i.e., $\logit(\mu_A) < 0$ in Figure 2.3. Hence, this member believes that his leader is more likely to have chosen low effort. When $Q_H$ is observed, the signal that the member receives now contradicts with his interim belief. However, he arrives at a posterior belief that is lower than that of a Bayesian (i.e., point $A''$ instead of point $A$). This implies that the member under-updates his belief when he receives a signal that contradicts with what he previously thought to be true. Another way to think about this is that a member who suffers from confirmatory bias actively looks for signals that confirm his interim belief, and therefore ignores any information that contradicts with this belief.

![Figure 2.4: Interpretation of $\gamma_G$ given $Q_H$ observed and $\delta = 1$](image)

Next, the parameters $\gamma_G$ and $\gamma_B$ allow us to capture the extent to which the group member responds to signals from the leader. Figure 2.4 shows the implications of different values of $\gamma_G$ on the relationship between the member’s posterior and interim beliefs.\(^{22}\)

Note that $\gamma_G, \gamma_B$ govern the intercepts of the regression conditional on the signal received by the member. If $\gamma_G > 1$, then the member is, on average, over-responsive to good

\(^{22}\)A similar analysis can be done for the case where a bad outcome is observed.
signals from the leader relative to a Bayesian, and tends to arrive at a posterior that is higher than that of a Bayesian. Specifically, the biased member attributes good outcomes of the leader more to effort as compared to an unbiased Bayesian member. On the other hand, if $\gamma_G < 1$, then the member is conservative in his response to good signals from the leader, and tends to arrive at a posterior that is lower than that of a Bayesian on average. In this case, the biased member attributes good outcomes of the leader more to luck as compared to an unbiased Bayesian member. Figure 2.4 also shows what happens when $\gamma_G = 0$ or $\gamma_G < 0$, which correspond to a non-updater and an inconsistent updater, respectively. We discuss these types of updating behavior in further detail in Section 2.6.3.

Finally, we can also capture asymmetric updating of beliefs, i.e., asymmetric attribution of outcomes to effort and luck. If $\gamma_G > \gamma_B$, then we say that the member is more (less) likely to attribute a good outcome from the leader to effort (luck). Conversely, if $\gamma_G < \gamma_B$, then we say that the member is more (less) likely to attribute a bad outcome from the leader to effort (luck).

### 2.6 Results

For the main analyses in this paper, we pool data from both the Game 0 and Game 1 treatments.\textsuperscript{23} For robustness, we show in Appendix A.4.1 that the main conclusions do not change qualitatively when we consider only the Game 1 treatments.

#### 2.6.1 The dictator game as a proxy for an individual’s type

We conjecture in Section 2.4 that subjects’ contributions in the dictator game provide us with a proxy of each individual’s type, i.e., subjects who transfer more of their endowment to their matched partner in the dictator game are more likely to choose $e_H$ when they are appointed as a leader. As our testable hypotheses depend on this relationship between the leader’s type and their effort choice, we first examine if it holds.

\textsuperscript{23}Hence, data from treatments LA0 and LA1 are pooled together as treatment LA, while data from treatments HA0 and HA1 are pooled together as treatment HA.
Figure 2.5 presents the distribution of subjects’ contributions in the dictator game against their effort choices across different appointment mechanisms. Because the subjects only participate in the dictator game once, the distribution of contributions are the same across different treatments. Within each panel in Figure 2.5, the black bars represent the proportion of leaders who choose high effort ($e_H$) while the gray bars represent the leaders who choose low effort ($e_L$).

A clear pattern that emerges is that leaders who are more altruistic in the dictator game are also the ones who are more likely to choose the investment option that is in the interest of the group (i.e., high effort). This pattern is consistent across the different appointment mechanisms.\footnote{Interestingly, there is a higher proportion of individuals who transferred nothing to their matched partner in the dictator game but who chose high effort as leaders in treatment HA compared to the other treatments. This may be because these individuals believe that they are unlikely to be appointed}
Table 2.2 presents marginal-effects estimates from a probit model for the relationship between the subjects' decisions as leaders in the leadership task and their dictator game behavior, both at the pooled level (column 1) and at the treatment level (columns 2 to 5). In the regression analyses, we control for order effects, the subjects’ behavior in the risk game, and Game 1.

The estimates in column (1) of Table 2.2 suggest that there exists a significantly positive relationship between the leader’s contribution in the dictator game and their decision to choose high effort in the leadership task. A leader who contributes 1% more to their matched partner in the dictator game is 0.4% more likely to choose $e_H$ in the leadership task on average, and this effect is statistically significant (p-value < 0.001). This result is also robust at the treatment level in columns (2) to (4), where a leader is between 0.3% and 0.5% significantly more likely to choose $e_H$ on average for every additional 1% of their endowment transferred in the dictator game.

Table 2.2: Regression of leader’s effort choice

<table>
<thead>
<tr>
<th></th>
<th>Pooled</th>
<th>RA</th>
<th>LA</th>
<th>HA</th>
<th>GA</th>
</tr>
</thead>
<tbody>
<tr>
<td>% endowment transferred in DG</td>
<td>0.004***</td>
<td>0.005***</td>
<td>0.005***</td>
<td>0.003**</td>
<td>0.004***</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(4.15)</td>
<td>(5.60)</td>
<td>(2.54)</td>
<td>(2.91)</td>
</tr>
<tr>
<td>% endowment invested in RG</td>
<td>-0.001</td>
<td>-0.002**</td>
<td>-0.000</td>
<td>-0.000</td>
<td>-0.001</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(2.40)</td>
<td>(0.51)</td>
<td>(0.48)</td>
<td>(0.90)</td>
</tr>
<tr>
<td>Game 1</td>
<td>-0.058***</td>
<td>-0.067*</td>
<td>-0.066</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.020)</td>
<td>(1.85)</td>
<td>(1.63)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Order Effects</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Observations</td>
<td>1,632</td>
<td>272</td>
<td>544</td>
<td>544</td>
<td>272</td>
</tr>
</tbody>
</table>

Marginal effects of probit model reported. Robust standard errors in parentheses. Standard errors are clustered at the subject level in column (1).

DG: Dictator Game; RG: Risk Game.

*** p<0.01, ** p<0.05, * p<0.10.

Moreover, we predict that all individuals will prefer to have the individual of the highest type ($\beta_i$) appointed as the leader. Figure 2.6 shows the subjects’ preferences for their leader’s type under treatment GA (panel a) and their beliefs about what the other to be the leader in this treatment, and therefore think that their effort choices will be less likely to be implemented.
members of their group have indicated as their preferences (panel b). The majority of the subjects (77.6%) prefer to have the most altruistic individual, i.e., the highest contributor in the dictator game, to be the leader. Similarly, the majority of the subjects (76.1%) believe that the other members of their group will prefer to have the highest contributor in the dictator game to be the leader.

Finally, we also elicit members’ beliefs about the leader’s behavior in the dictator game under each appointment mechanism. Figure 2.7 shows the average interim belief reported by group members, separated by their belief about how much the leader appointed in a given round has transferred in the dictator game. The graph pools the data across all
treatments. A clear pattern that emerges is that members’ interim beliefs increase on average when they believe that the leader has transferred more of their endowment in the dictator game. Hence, we find support that the subjects also regard the dictator game as a predictor of how likely an individual will act in the group’s interest by choosing high effort as a leader.

Overall, we conclude that the dictator game is a good proxy for an individual’s type $\beta_i$.

### 2.6.2 Members’ interim beliefs

We now examine the members’ interim beliefs after they have observed the appointment mechanism, but prior to observing their leader’s outcomes. In all of our analyses, belief is a variable that takes an integer value in $[0, 100]$, where a higher belief implies that the member thinks the leader is more likely to have chosen high effort ($e_H$). Figure 2.8 presents the distributions of interim beliefs of the members by treatment. In each panel, the dashed line represents the mean interim belief.

The histograms in Figure 2.8 suggest that the group members respond to the mechanism used to select their leader. When the leader is randomly assigned in treatment RA, the members’ beliefs are approximately centered on 50% (panel a). The mean interim belief in treatment RA is 45.94%.

The distribution of interim beliefs is highly skewed to the right in treatment LA where the most selfish individual in the group is assigned to be the leader (panel b), and this is statistically significantly different to that in treatment RA (Kolmogorov-Smirnov test: $p$-value < 0.001). Under this appointment mechanism, members strongly believe that their leader will likely act in self-interest and not in the interest of the group. The members’ average interim belief is also the lowest in treatment LA, at 34.15%, and a Wilcoxon signed-rank test rejects the null hypothesis that the members’ interim beliefs are equal between treatments RA and LA ($p$-value < 0.001).

On the other hand, the distribution of beliefs is skewed to the left in treatment HA where the most altruistic individual is appointed to be the leader of the group (panel
c). This is statistically significantly different from the distribution of beliefs in treatment RA (Kolmogorov-Smirnov test: $p$-value < 0.001). The members’ average interim belief is the highest in this treatment, at 57.40%. A Wilcoxon signed-rank test rejects the null hypothesis that the members’ interim beliefs are equal between these two treatments ($p$-value < 0.001).

When the leader is appointed based on the preferences of the group in treatment GA (panel d), the distribution of interim beliefs shifts slightly to the right relative to that in treatment RA, but the distributions of beliefs between these two treatments are not statistically significantly different (Kolmogorov-Smirnov test: $p$-value = 0.664). Nonetheless, the average interim belief in treatment GA is slightly higher than that in treatment RA, at 48.65%. Using a Wilcoxon signed-rank test, we reject the null
hypothesis that the members’ interim beliefs are equal between these two treatments (p-value = 0.006).

Table 2.3: Regression of members’ interim belief

<table>
<thead>
<tr>
<th>Variables</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1.417)</td>
<td>(1.416)</td>
<td>(1.403)</td>
<td>(1.372)</td>
</tr>
<tr>
<td>Treatment HA</td>
<td>9.982***</td>
<td>9.982***</td>
<td>8.950***</td>
<td>9.359***</td>
</tr>
<tr>
<td></td>
<td>(1.311)</td>
<td>(1.309)</td>
<td>(1.263)</td>
<td>(1.246)</td>
</tr>
<tr>
<td>Treatment GA</td>
<td>2.717**</td>
<td>2.717**</td>
<td>1.857</td>
<td>2.198*</td>
</tr>
<tr>
<td></td>
<td>(1.355)</td>
<td>(1.353)</td>
<td>(1.273)</td>
<td>(1.271)</td>
</tr>
<tr>
<td>Chooses high effort as leader</td>
<td></td>
<td></td>
<td>23.382***</td>
<td>14.109***</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(1.848)</td>
<td>(1.588)</td>
</tr>
<tr>
<td>% endowment invested in RG</td>
<td>-0.104**</td>
<td>-0.073**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.043)</td>
<td>(0.036)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Game 1</td>
<td>-2.952***</td>
<td>-2.952***</td>
<td>-1.405</td>
<td>-2.018**</td>
</tr>
<tr>
<td></td>
<td>(1.041)</td>
<td>(1.040)</td>
<td>(0.955)</td>
<td>(0.964)</td>
</tr>
<tr>
<td>Constant</td>
<td>59.182***</td>
<td>48.890***</td>
<td>48.117***</td>
<td>44.066***</td>
</tr>
<tr>
<td></td>
<td>(3.916)</td>
<td>(1.237)</td>
<td>(3.534)</td>
<td>(1.243)</td>
</tr>
<tr>
<td>Order Effects</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>Individual FE</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>Observations</td>
<td>1.632</td>
<td>1.632</td>
<td>1.632</td>
<td>1.632</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.137</td>
<td>0.251</td>
<td>0.286</td>
<td>0.305</td>
</tr>
</tbody>
</table>

Robust standard errors clustered at the subject level in parentheses. For all regressions, treatment RA is the reference treatment.
RG: Risk Game.
*** p<0.01, ** p<0.05, * p<0.10.

These results are supported by regression analyses. Table 2.3 presents OLS estimates for the regressions of interim beliefs against treatment variables. In the analyses, we control for Game 1, order effects (in columns 1 and 3), and individual fixed effects (in columns 2 and 4). In all the specifications, treatment RA is the comparison group. The coefficient estimates for the treatment variables in columns (1) and (2) support our conclusions from the non-parametric analysis. In the regression analysis, we can also control for the subjects’ own decision as a leader (columns 3 and 4). The coefficient estimates suggest that the treatment effects remain identical both in direction and magnitude, although the estimates for treatment GA are now statistically insignificant in column (3).
and marginally statistically significant in column (4) (p-values = 0.159 and 0.093, respectively). Moreover, a subject who chooses to exert high effort as a leader under a specific appointment mechanism is also more likely to, as a group member, expect their leader to choose high effort under the same appointment mechanism. Overall, we find support for Hypothesis 2.1, and we summarize our results as follows.

Result 2.1: Group members respond to the leadership appointment mechanism in their interim beliefs. Compared to when the leader is randomly assigned, the interim beliefs are statistically significantly lower (higher) on average when the most selfish (altruistic) individual is appointed to be the leader. Hence, $\mu_{LA}^i < \mu_{RA}^i < \mu_{HA}^i$. The members’ interim beliefs are statistically significantly higher on average when the leader is appointed by the group than under random assignment, i.e., $\mu_{RA}^i < \mu_{GA}^i$.

2.6.3 Group members’ posterior beliefs

Figure 2.9 presents a summary of the members’ interim and posterior beliefs, conditional on observing a good outcome (panel a) and a bad outcome (panel b) from the leader, respectively. Within each panel, the graph on the left shows the distributions of both the interim and posterior beliefs for all the subjects across all six rounds in the leadership task. The histograms show that, on average, the members’ posterior beliefs move in the same direction as the observed signal from the leader. At the pooled level, the members’ interim belief that the leader has chosen $e_H$ is 46.28% on average. The distribution of posterior beliefs is skewed to the left and members’ beliefs increase to 56.72% on average when a good outcome is observed. On the other hand, the distribution of posterior beliefs is skewed to the right and members’ beliefs decrease to 33.35% on average when a bad outcome is observed.

---

This result may be driven by a false consensus effect, i.e., subjects tend to overestimate the extent to which other individuals behave or think in a similar way to themselves (e.g., see Ross et al., 1977; Marks and Miller, 1987). Interestingly, we also find a positive correlation between the subjects’ own behavior in the dictator game and their beliefs about the leader’s behavior in the dictator game (p-value < 0.001 at the pooled level), which is also consistent with a false-consensus effect. Moreover, while the subjects’ behavior in the dictator game has no statistically significant impact on their interim belief about the leader’s effort choice (p-value = 0.512), their beliefs about the leader’s dictator game behavior have a positive impact on their interim beliefs (p-value < 0.001).
Within each panel in Figure 2.9, the graph on the right shows the members’ posterior beliefs given their interim beliefs at the subject-round level, where the size of each bubble is proportional to the frequency of each pair of interim/posterior beliefs. The solid line represents the posterior beliefs derived under Bayes’ rule given the members’ interim beliefs, while the dashed line represents posterior beliefs that are equal to interim beliefs. The bubble plots reveal that the group members significantly deviate from the Bayesian benchmark when revising their beliefs. A large proportion of members under-update their beliefs relative to a Bayesian, i.e., by having posterior beliefs between the solid and dashed lines in the figure. However, a modest proportion of members over-update their beliefs, where their beliefs are above (below) that predicted by Bayes’ rule when a good (bad) outcome is observed. Moreover, the bubble plots reveal that a non-trivial proportion of members either: (i) do not tend to revise their beliefs at all, i.e., the bubbles are on the
dashed line; or (ii) revise their beliefs in a direction opposite to the observed signal, i.e.,
the bubbles are below (above) the dashed line when a good (bad) outcome is observed. We discuss these two types of updating behavior in detail in the next section.

**Inconsistent and non-updaters**

Figure 2.10 presents the distribution of subjects based on the number of inconsistent and non-updates throughout the experiment. A belief update is classified to be *inconsistent* if the reported posterior belief is in a direction opposite to the observed state (e.g., the member revises their belief upwards when a bad outcome is observed, or vice versa). On the other hand, a *non-update* is one where the member’s posterior belief is equal to their interim belief.

The histograms reveal that a non-trivial proportion of subjects tend to persistently report posterior beliefs in a direction opposite to the observed outcome and/or posterior beliefs that are equal to their interim beliefs. Despite the detailed instructions and having a practice round, it is possible that some subjects may fail to understand the experiment or fully engage with the task during the actual experiment. Specifically, inconsistent updating of beliefs may arise if subjects do not fully understand the experiment, while non-updating may be due to subjects failing to engage with the task during the experiment. The inclusion of these observations may bias the interpretation of the results,
particularly if these subjects are reporting beliefs that do not genuinely reflect their true posterior beliefs. Hence, for the remainder of the analysis, we exclude 44 (out of 272) subjects where 25% or more of their posterior beliefs are inconsistent and another 23 subjects who report posterior beliefs equal to their interim beliefs across all six rounds of the experiment. These two groups jointly constitute 24.6% of the sample. These numbers are largely in line with what is found in the existing literature (see, e.g., Möbius et al., 2014; Barron, 2016; Coutts, 2017).

**Comparison against Bayesian posteriors**

We first examine the members’ average responsiveness to the leader’s outcomes vis-à-vis the Bayesian benchmark. For each interim belief reported by the group members, we calculate what their implied posterior belief would be under Bayes’ rule. Figure 2.11 presents the actual posterior beliefs of the group members versus the Bayesian posteriors as implied by their interim beliefs, conditional on observing a good outcome (panel a) and a bad outcome (panel b), respectively. The size of each bubble is proportional to the number of observations.

![Figure 2.11: Actual vs. Bayesian posterior beliefs (less inconsistent and non-updaters)](image)

A clear pattern that emerges from panel (a) of Figure 2.11 is that, conditional on

---

26 We find that the inclusion of these subjects leads to an attenuation of the coefficient estimates of $\gamma_G$ and $\gamma_B$, which is similar to what Barron (2016) has observed. Nonetheless, most of the results remain the same qualitatively. For robustness, we present the analyses including these subjects in Appendix A.4.2.
observing a good outcome, members tend to report posterior beliefs that are lower than the Bayesian benchmark. Majority of the observations lie below the 45-degree line. This suggests that they are, on average, revising their beliefs (upwards) less than a Bayesian would when they observe a good outcome.

On the other hand, panel (b) reveals that, while members tend to report posterior beliefs that are slightly higher than the Bayesian benchmark on average conditional on observing a bad outcome, many of them also have posterior beliefs that lie below the 45-degree line. As a result, there is no clear indication as to which direction deviations from the Bayesian posteriors occur systematically. It is likely that, on average, members tend to report posterior beliefs that are closer to the Bayesian benchmark when they observe a bad outcome.

Overall, Figure 2.11 suggests that there is an asymmetry in the way members treat good and bad outcomes. In the next section, we examine this issue further by analyzing the specific channels through which deviations from Bayes’ rule occur.

**Estimating different channels of deviations from Bayes’ rule**

We now estimate equation (2.7) using ordinary least squares (OLS) to analyze the biases that members suffer from when updating their beliefs.\(^ {27}\) Table 2.4 presents the regression results both at the pooled level (column 1) and at the treatment level (columns 2 to 5).\(^ {28}\) As a test of Hypothesis 2.2, our primary interest is to examine whether the coefficients are different from 1. Hence, the asterisks next to each coefficient estimate denote the significance of a t-test of whether that coefficient is statistically different from 1. The last two rows of Table 2.4 present the results of a Wald test of equality between \(\gamma_G\) and \(\gamma_B\).

\(^ {27}\)One concern with estimating (2.7) using OLS is that the estimates are biased if there are measurement errors in the subjects’ reported beliefs. For example, subjects could make mistakes or are imprecise when reporting their beliefs. For robustness, we also consider an alternative specification where the appointment mechanisms are used as instruments for the logit of members’ interim beliefs for the analysis at the pooled level. This instrumental-variable (IV) approach assumes that the appointment mechanisms have a direct effect on the members’ interim beliefs (as we have shown to be true in Section 2.6.2) but not on their posterior beliefs. We find that the IV estimates lead to similar conclusions. Details of the results from the IV regression can be found in Appendix A.4.3.

\(^ {28}\)Note that the logit function is only defined for beliefs in \((0,100)\). Instead of excluding observations of subjects who state 0 or 100 as their interim or posterior belief about the leader, if a subject indicates a belief of 0 (100), we take the logit of 0.01 (99.99) as an approximation.
giving us a test of the presence of an asymmetric attribution bias.

Table 2.4: Regression of members’ posterior beliefs

<table>
<thead>
<tr>
<th>Variables</th>
<th>(1) Pooled</th>
<th>(2) RA</th>
<th>(3) LA</th>
<th>(4) HA</th>
<th>(5) GA</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \delta ): Logit(interim belief)</td>
<td>0.695***</td>
<td>0.764***</td>
<td>0.692***</td>
<td>0.703***</td>
<td>0.529***</td>
</tr>
<tr>
<td></td>
<td>(0.039)</td>
<td>(0.071)</td>
<td>(0.054)</td>
<td>(0.058)</td>
<td>(0.135)</td>
</tr>
<tr>
<td>( \gamma_G ): Good outcome ( \times ) logit(( p ))</td>
<td>0.751***</td>
<td>0.744***</td>
<td>0.622***</td>
<td>0.847*</td>
<td>0.798**</td>
</tr>
<tr>
<td></td>
<td>(0.051)</td>
<td>(0.089)</td>
<td>(0.079)</td>
<td>(0.081)</td>
<td>(0.098)</td>
</tr>
<tr>
<td>( \gamma_B ): Bad outcome ( \times ) logit(( 1 - p ))</td>
<td>0.966</td>
<td>0.932</td>
<td>1.058</td>
<td>0.946</td>
<td>0.876</td>
</tr>
<tr>
<td></td>
<td>(0.067)</td>
<td>(0.092)</td>
<td>(0.117)</td>
<td>(0.072)</td>
<td>(0.114)</td>
</tr>
<tr>
<td>Observations</td>
<td>2,460</td>
<td>410</td>
<td>820</td>
<td>820</td>
<td>410</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.608</td>
<td>0.686</td>
<td>0.651</td>
<td>0.583</td>
<td>0.421</td>
</tr>
</tbody>
</table>

Test of \( \gamma_G = \gamma_B \)

| test statistic | -3.190 | -1.588 | -3.065 | -1.081 | -0.512 |
| p-value        | 0.002  | 0.114  | 0.002  | 0.281  | 0.609  |

Robust standard errors clustered at the subject level in parentheses. This analysis excludes subjects classified as inconsistent or non-updaters.

*** p<0.01, ** p<0.05, * p<0.10. Null hypothesis is coefficient = 1.

Column (1) shows that members are biased in their belief-updating process. The estimate for \( \delta \) suggests that members suffer from base-rate neglect (i.e., \( \delta < 1 \)). Relative to a Bayesian, members place too little weight on prior information on average (test of \( \delta = 1 \): p-value < 0.001). At the same time, the estimate for \( \gamma_G \) suggests that members tend to attribute good outcomes to luck more than a Bayesian would (i.e., \( \gamma_G < 1 \)), and this effect is statistically significant (test of \( \gamma_G = 1 \): p-value < 0.001). However, there is no statistically significant evidence that members attribute bad outcomes to the leader’s effort differently from a Bayesian (test of \( \gamma_B = 1 \): p-value = 0.608). Hence, relative to the Bayesian benchmark, group members give too little credit for their leader’s success but apportion enough blame for their leader’s failure. A Wald test of \( \gamma_G = \gamma_B \) also reveals that, overall, members update their beliefs about the leader asymmetrically (i.e., \( \gamma_G < \gamma_B \)). Members tend to attribute good outcomes more to luck and, relatively, bad outcomes more to the leader’s effort. This effect is statistically significant (p-value = 0.002).
Overall, we fail to find support for Hypothesis 2.2(i) in that members are not Bayesian when updating their beliefs after observing their leader’s outcomes. We summarize our findings as follows.

**Result 2.2:** Group members suffer from base-rate neglect in their updating behavior, i.e., $\delta < 1$. Compared to the Bayesian benchmark, members attribute good outcomes more to the leader’s luck, i.e., $\gamma_G < 1$, but they are no different from a Bayesian when attributing bad outcomes to the leader’s effort and luck, i.e., $\gamma_B = 1$. Overall, members attribute good (bad) outcomes more to the leader’s luck (effort), i.e., $\gamma_G < \gamma_B$.

We now analyze the members’ updating behavior across the different appointment mechanisms (Hypothesis 2.2(ii)). The coefficient estimates in columns (2) to (5) reveal that we also observe similar biases in updating behavior at the treatment level as compared to the pooled level. Under each appointment mechanism, members consistently suffer from base-rate neglect (i.e., $\delta < 1$), attribute good outcomes of the leader more to luck as compared to a Bayesian would (i.e., $\gamma_G < 0$), but attribute bad outcomes no differently from a Bayesian (i.e., $\gamma_B = 1$). Moreover, we find that members attribute good (bad) outcomes more to luck (effort), i.e., $\gamma_G < \gamma_B$, only in treatment LA (Wald test of $\gamma_G = \gamma_B$: p-value = 0.002), but not in treatments RA, HA, and GA (Wald tests of $\gamma_G = \gamma_B$: p-values = 0.114, 0.281 and 0.609, respectively). Nonetheless, the direction of the asymmetry in the treatment of outcomes in treatments RA, HA, and GA is still similar to that of treatment LA and at the pooled level.\(^{29}\)

Comparing the magnitudes of the biases across the appointment mechanisms, we fail to reject the null hypotheses that the estimates for $\delta$, $\gamma_G$, and $\gamma_B$ are jointly equal to one another (Wald tests of $\delta^{RA} = \delta^{LA} = \delta^{HA} = \delta^{GA}$: p-value = 0.395; $\gamma_G^{RA} = \gamma_G^{LA} = \gamma_G^{HA} = \gamma_G^{GA}$: p-value = 0.110; and $\gamma_B^{RA} = \gamma_B^{LA} = \gamma_B^{HA} = \gamma_B^{GA}$: p-value = 0.686). However, we observe that the coefficient estimate for $\gamma_G$ in treatment LA is lower than that in treatments HA and GA, while the estimate for $\gamma_B$ is higher in treatment LA than that in treatment GA. Pairwise comparisons reveal that the coefficient estimate for $\gamma_G$ in

\(^{29}\) The reduction in the statistical significance of these results may be because there are fewer observations for the analyses at the treatment level.
treatment LA is statistically significantly lower than that in treatment HA (p-value = 0.020), and marginally statistically significantly lower than that in treatment GA (p-value = 0.099). However, the difference in the coefficient estimate for $\gamma_B$ is not statistically significantly different between treatments LA and GA (p-value = 0.232). Hence, members attribute good outcomes of the leader even more to luck than a Bayesian would when the most selfish individual is appointed to be the leader as compared to when the most altruistic individual is appointed to be the leader or when the leader is appointed by the group.

Moreover, pairwise comparisons of $(\gamma_G - \gamma_B)$ also reveal that the way members attribute good vs. bad outcomes is statistically significantly different between treatments LA and HA (p-value = 0.039) and LA and GA (p-value = 0.077). This difference appears to be largely driven by the fact that, as noted above, members are more biased in their attribution of good outcomes of the leader (to luck) when the most selfish individual is appointed to be the leader.

Overall, we do not find support for Hypothesis 2.2(ii). Group members do not behave like Bayesian agents under each appointment mechanism. They tend to exhibit similar biases in their updating behavior across different appointment mechanisms. However, we find that members tend to attribute good (bad) outcomes more to luck (effort) only when the most selfish individual is appointed to be the leader.

**Result 2.3:** Group members are biased in their updating behavior under all the appointment mechanisms, but they exhibit largely similar biases regardless of the way the leader has been appointed. Across all the appointment mechanisms, members suffer from base-rate neglect ($\delta < 1$), attribute good outcomes more to luck than a Bayesian would ($\gamma_G < 1$), and are no different from a Bayesian when attributing bad outcomes to the leader’s effort and luck ($\gamma_B = 1$). However, members attribute good (bad) outcomes more to luck (effort) only in treatment LA but not in treatments RA, HA, and GA.
**Heterogeneity in updating behavior**

As revealed in Figure 2.9, members appear to be heterogeneous in their updating behavior. The biases and asymmetry in updating behavior that we find at the aggregate level may not necessarily hold once we consider heterogeneity in updating behavior. To explore this further, Table 2.5 presents the results from a 2-component finite mixture model analysis of members’ updating behavior at the pooled level.\(^{30}\)

**Table 2.5: Finite mixture model for updating behavior**

<table>
<thead>
<tr>
<th>Variables</th>
<th>Dependent variable: Logit(posterior)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Component 1</td>
</tr>
<tr>
<td>(\delta): Logit(interim belief)</td>
<td>0.936***</td>
</tr>
<tr>
<td></td>
<td>(0.011)</td>
</tr>
<tr>
<td>(\gamma_G): Good outcome (\times) logit((p))</td>
<td>0.535***</td>
</tr>
<tr>
<td></td>
<td>(0.031)</td>
</tr>
<tr>
<td>(\gamma_B): Bad outcome (\times) logit(1−(p))</td>
<td>0.668***</td>
</tr>
<tr>
<td></td>
<td>(0.044)</td>
</tr>
<tr>
<td>Test of (\gamma_G = \gamma_B)</td>
<td>-3.47</td>
</tr>
<tr>
<td>p-value</td>
<td>0.001</td>
</tr>
<tr>
<td>Latent Class Marginal Probabilities</td>
<td></td>
</tr>
<tr>
<td>(\mu)</td>
<td>0.889</td>
</tr>
<tr>
<td></td>
<td>(0.020)</td>
</tr>
<tr>
<td>Model Fit</td>
<td></td>
</tr>
<tr>
<td>Log likelihood</td>
<td>-3317.86</td>
</tr>
<tr>
<td>AIC</td>
<td>6653.72</td>
</tr>
<tr>
<td>BIC</td>
<td>6705.99</td>
</tr>
</tbody>
</table>

Robust standard errors clustered at the subject level in parentheses. This analysis excludes subjects classified as inconsistent or non-updaters.

*** \(p<0.01\), ** \(p<0.05\), * \(p<0.10\). Null hypothesis is coefficient = 1.

We also consider 3-component and 4-component finite mixture models. We make the following considerations when deciding the number of latent classes to include in the finite mixture model analysis. Moving from a 2-component to a 3-component model results in a modest decrease in the AIC (BIC) from 6653.72 (6705.99) to 6084.22 (6165.53), while moving from a 3-component model to a 4-component model leads to an increase in the AIC and BIC to 6432.80 and 6543.16, respectively. Hence, the AIC and BIC would have guided us towards choosing a 3-component model. However, moving to a 3-component model results in a class of updating behavior characterized by a negative coefficient for \(\delta\), and this class constitutes only 4.8% of the sample. This suggests that the additional component is picking up noisy updating behavior of the members (despite having dropped some of the inconsistent and non-updaters). Given these considerations, we decided to proceed with a 2-component finite mixture model analysis.
The coefficient estimates in Table 2.5 suggest that the majority of updates in the sample (component 1: 88.9%) suffer from base-rate neglect although the magnitude of this bias is small ($\delta = 0.936$, p-value < 0.001). Moreover, on average, this group of belief updates is characterized by the attribution of the leader’s outcomes more to luck as compared to a Bayesian would ($\gamma_G = 0.535$ and $\gamma_B = 0.668$ with p-values < 0.001 for both coefficients). Within these belief updates, members also tend to attribute good outcomes more to luck and bad outcomes more to effort ($\gamma_G < \gamma_B$), and this effect is statistically significant (p-value = 0.001).

The second group of updates (component 2) constitutes about 11.1% of the sample. This group of updates can be characterized as suffering from strong base-rate neglect ($\delta = 0.148$, p-value < 0.001) and an over-responsiveness to the outcomes of the leaders on average ($\gamma_G = 1.936$ and $\gamma_B = 1.945$ with p-values = 0.021 and 0.020, respectively). Within this group of updates, the group members strongly neglect their interim beliefs. Moreover, on average, this group of updates is characterized by an attribution of the leader’s outcomes more to effort as compared to a Bayesian would. However, these updates do not suffer from any asymmetric bias in the attribution of their leader’s outcomes (test of $\gamma_G = \gamma_B$: p-value = 0.984).

Overall, our finite mixture model analysis suggests that there is heterogeneity in the members’ updating behavior, and that the majority of the belief updates in the sample is characterized by a modest level of base-rate neglect, under-responsiveness to the leader’s outcomes, as well as an asymmetric attribution of the leader’s outcomes to effort and luck.

2.7 Conclusion

In many environments, the determinants of outcomes are not observable. What beliefs do individuals hold in such circumstances about the role of luck versus effort in the determination of outcomes? Do these beliefs depend on the outcome, i.e., whether the outcome is good or bad? Focusing on leadership, these are the questions we address in
In particular, this paper studies group members’ beliefs about the decisions of their leaders when effort choices of leaders are not observable. We find that the leadership appointment mechanism is important in that it affects the expectations (unconditional beliefs) members have about the type of their leader. We also find that after observing the outcome of the leader’s investment decision, members suffer from biases in their belief-updating process. As compared to a Bayesian, they attribute good outcomes more to luck. However, in their response to bad outcomes, they are no different from a Bayesian. The asymmetry we observe in the way good and bad outcomes are treated suggests that the credit leaders receive for good outcomes is less than the blame they get for bad outcomes. Moreover, relative to the Bayesian benchmark, members suffer from base-rate neglect in that in the belief updating process, they place too little weight on their beliefs prior to the realization of the outcome. Importantly, we find that the biases in updating behavior tend to be driven by those subjects who choose low effort as leaders.

Our study is a first step in understanding biases in the evaluation of decision-makers’ choices. We consider a set-up where the group members’ beliefs do not have an impact on the compensation received by the leader. This design choice has the advantage of allowing us to study biases in an environment not confounded by factors such as higher-order beliefs and social preferences. In practice, leadership is typically endogenous, and leaders and their group members are likely to engage in repeated interactions. Hence, our study is an essential first step for studying biases in such environments where beliefs are likely to be payoff-relevant to the leaders. If beliefs are payoff-relevant, considerations such as social preferences and higher-order beliefs may have an impact on behavior. Whether biases would be different in such situations is left for future research.
Chapter 3

Eliciting beliefs about actions of others: A comparison between the quadratic and binarized scoring rules

3.1 Introduction

The behavior of individuals is often driven by the beliefs they hold, be it about their own ability, the actions of others, or even nature. For example, a potential buyer’s decision to purchase an item that is currently on sale depends on whether they think the store would offer a further discount the next day. Being able to know what beliefs individuals hold can help improve our understanding of the underlying motivations of decision-makers. As a result, eliciting the beliefs of individuals is becoming increasingly common in experimental economics, where these beliefs are often expressed in the form of probabilistic statements, i.e., subjective probabilities.

However, getting subjects to truthfully report these subjective probabilities can be challenging for a number of reasons. First, in line with the methodologies used in experimental economics, the mechanism used to elicit subjects' beliefs must be incentive compatible in that it encourages the subjects to truthfully report their underlying beliefs. Second, the mechanism or scoring rule should not be too complex, as otherwise it may generate noisy belief reports as a result of confusion among the subjects.

Given these concerns, researchers have developed various mechanisms to elicit prob-
abilistic beliefs from experimental subjects. One class of mechanisms, known as “proper scoring rules”, allows the experimenter to elicit truthful reports of the subject’s beliefs under the assumption of risk neutrality (see, e.g., Savage, 1971; Schervish, 1989). Examples include the quadratic scoring rule, the logarithmic scoring rule, and the spherical scoring rule. Among these, the quadratic scoring rule (QSR) is one of the more popular and commonly used methods, mainly because it is relatively intuitive and simple for subjects to understand.

Nonetheless, a major drawback of the QSR is that it incentivizes truth-telling only under the assumptions of risk neutrality and subjective expected utility maximization. When eliciting beliefs over binary events, theory predicts that risk-averse subjects have an incentive to distort their reports under the QSR. Under the assumption of subjective expected utility maximization, it is optimal for risk-averse subjects to distort their reported subjective probabilities towards the midpoint (i.e., 0.5). Recent studies have found evidence of such distortions in laboratory experiments (e.g., see Offerman et al., 2009; Hossain and Okui, 2013; Harrison et al., 2014). Biases in the subjects’ belief reports can pose a problem to experimenters hoping to draw inferences about the subjects’ underlying behavior. This is particularly worrying if the biases are correlated with preferences (i.e., over risks) that the researchers may also be interested in studying as part of their research agenda.

In this paper, we evaluate the performance of the QSR and compare it against the binarized scoring rule (BSR), which is a binary lottery procedure designed to induce risk neutrality in belief elicitation tasks (Hossain and Okui, 2013). The implementation of the BSR is similar to that of the QSR. Under both mechanisms, subjects’ payoffs are determined using a loss function which penalizes the subjects based on the square of the distance of their reported belief from the true state of the world. The key difference is that the subject’s payoff is directly penalized by this loss function under the QSR, while it is the probability of receiving a fixed payoff that is penalized under the BSR. Under the assumption of expected subjective utility maximization, the BSR is incentive compatible independent of the subjects’ risk preferences. However, this is not true for the QSR.
We are interested in the BSR because, while it is becoming increasingly common in experimental economics (e.g., see Hoffman, 2016; Babcock et al., 2017), few studies have evaluated its performance in the laboratory. Moreover, while the theoretical properties are different between the two scoring rules, the implementation of the BSR is similar to that of the QSR. Hence, we expect there to be minimal differences in the subjects’ ability to comprehend the two mechanisms.\footnote{Differences in the subjects’ abilities to understand the mechanisms may also induce different behavioral responses to each mechanism.} This can allow us to better understand the impact that subjects’ risk preferences may have on the beliefs elicited under the QSR.

We compare the use of the QSR and the BSR in an environment where subjects are asked to report their beliefs about the decisions made by others under risk. These decisions are payoff-relevant to the observers and may be affected by other-regarding preferences of the decision-maker. Specifically, this paper uses the same task as Chapter 2. Subjects participate in a decision task in groups of three where a leader makes an unobservable binary investment decision on behalf of the group. Group members are then asked to report their beliefs about the leader’s decision. We elicit both the members’ interim beliefs and their posterior beliefs conditional on observing the outcome of the leader’s investment decision. We implement a between-subject treatment design, where the main treatment variable of interest is the scoring rule used to incentivize the group members’ belief reports.

We examine two issues that may arise under the QSR when subjects exhibit risk-averse behavior. First, we ask whether the QSR leads to distortions in subjects’ reports of their interim beliefs. Assuming subjective expected utility maximization, theory predicts that it is optimal for risk-averse subjects to distort their reported beliefs towards the midpoint (i.e., 0.5) under the QSR. This distortion does not occur under the BSR. Hence, we compare the members’ reports of their interim beliefs between the QSR and BSR treatments and evaluate whether the QSR leads to belief reports that are closer to the midpoint than the BSR on average. Second, we examine whether the biases in updating behavior (i.e., deviations from Bayes’ rule) are different between the two treatments.

Overall, we do not find any statistically significant differences in the interim beliefs
reported by the group members between the QSR and BSR treatments. Contrary to our theoretical prediction, there is no statistically significant evidence that the interim beliefs elicited under the QSR are closer to the midpoint on average as compared to those elicited under the BSR. Moreover, we do not find any systematic differences in the updating behavior between the QSR and BSR treatments. Specifically, using the data obtained under the BSR in Chapter 2, we evaluate the belief reports and biases that would have been observed under the QSR. We find that the few differences in the actual updating behavior between the QSR and BSR treatments are inconsistent with our predictions.

The paper proceeds as follows. We describe our contributions to the related literature in the next section. We present the experimental design in Section 3.3, and in Section 3.4 we discuss our theoretical predictions. The results are presented in Section 3.5. Finally, Section 3.6 suggests avenues for future work by discussing possible reasons for the differences between our results and theoretical predictions.

### 3.2 Related Literature

Over the years, experimental economists have used various mechanisms to elicit subjects’ beliefs, both in the laboratory and in the field. In this paper, we focus on mechanisms that are used to elicit subjective probabilities over a single event (i.e., where there are binary states of the world).\(^2\) When eliciting subjective probabilities, researchers often employ a “scoring rule”, which is a function that maps both the subjects’ belief reports and the true state of the world (about which beliefs are elicited) to their experimental payment. A popular class of scoring rules, known as “proper scoring rules”, incentivize subjects to reveal their true subjective probabilities under the assumption that they are risk neutral (Savage, 1971; Schervish, 1989). Examples of proper scoring rules include the QSR, the logarithmic scoring rule, and the spherical scoring rule.

However, one concern regarding the use of proper scoring rules such as the QSR is

\(^2\)Both Schotter and Trevino (2014) and Schlag et al. (2015) provide extensive reviews of the mechanisms used to elicit beliefs across different domains in the literature.
that the assumption that subjects exhibit risk-neutral behavior is often violated. There is abundant evidence in the literature to suggest that subjects in the laboratory exhibit behavior consistent with risk aversion (e.g., Gneezy and Potters, 1997; Holt and Laury, 2002; Eckel and Grossman, 2002, 2008). To address this issue, Offerman et al. (2009) and Andersen et al. (2014) propose methods to directly correct the belief reports elicited under the QSR and recover the subjects’ underlying subjective probabilities, e.g., by using additional belief reports or decisions made under risk.

More recently, researchers have relied on alternative scoring rules that are robust to subjects’ risk preferences. One example is the “stochastic” variation of the Becker-DeGroot-Marschak (BDM) mechanism (Becker et al., 1964). This mechanism elicits the subjects’ reservation probabilities by asking them to indicate the point at which they are indifferent between two lotteries. For the mechanism to be incentive compatible, one only needs to rely on the assumption that subjects satisfy stochastic dominance (Karni, 2009). However, implementing the stochastic BDM mechanism is challenging as it can often be too complicated for the subjects to understand. As a result, several studies have considered different ways of implementing the mechanism in the laboratory (e.g., see Holt and Smith, 2009; Hao and Houser, 2012; Trautmann and van der Kuilen, 2015). An alternative mechanism, which we examine in this paper, is the BSR. This mechanism is a simple modification of the QSR in that subjects are rewarded for their beliefs probabilistically instead of deterministically. With this modification, the scoring rule is incentive compatible independent of the subjects’ risk preferences (Hossain and Okui, 2013; Harrison et al., 2014).³

Given the various belief-elicitation mechanisms available in the literature, there is growing interest among researchers in evaluating the ability of these mechanisms to elicit subjects’ beliefs reliably. We contribute to this literature by comparing the QSR and the BSR within an environment where individuals are asked to state their beliefs about the decisions of others. While the majority of the studies in the literature have compared the QSR against other proper scoring rules (such as the logarithmic scoring rule) or

³Note that Harrison et al. (2014) refer to the BSR as the quadratic scoring rule modified using a binary lottery procedure.
elicitation mechanisms (such as certainty equivalent or unincentivized beliefs), few have
directly compared the QSR with the BSR.\footnote{In this section, we focus our discussion on the papers that directly compare the QSR and the BSR. The interested reader may refer to Schlag et al. (2015), who present a detailed summary of the studies that empirically compares various other elicitation mechanisms.}

Two studies that are most closely related to ours are Hossain and Okui (2013) (HO) and Harrison et al. (2014) (HMS), who also compare the QSR and the BSR. Both of these studies find that the BSR performs better than the QSR, in that the QSR leads to more distorted and less accurate belief reports as compared to the BSR. On the one hand, HO induce the subjects’ beliefs by directly providing them with objective probabilities of the states of the world. They evaluate the scoring rules by comparing the subjects’ belief reports against these objective probabilities.\footnote{This is a common approach used by researchers to evaluate the performance of elicitation mechanisms given that the underlying beliefs of the subjects are unobservable. See, e.g., Armantier and Treich (2013) and Offerman and Palley (2016) who also use this approach to compare other types of elicitation mechanisms.} They find that the BSR leads to belief reports that are closer to the true objective probabilities than the QSR. On the other hand, HMS compare the QSR and the BSR in an environment where the objective probabilities are not known to the subjects. Subjects are asked to state their beliefs about the composition of a bingo cage after they have observed the cage spun a number of times. They find that the QSR leads to reported beliefs that are both closer to the midpoint and less accurate in predicting the composition of the bingo cage.

However, we differ from these two studies on several aspects. First, unlike HO, we evaluate the performance of the scoring rules in an environment where the objective probability of the underlying event (in our case, the leader’s investment decision) is not known to the subjects. While this is also the case in HMS, a key feature of our design is that we consider an environment where subjects are asked to state their beliefs about the decisions made by others in a strategic environment. This is important because researchers are often interested in understanding subjects’ beliefs about the strategies or actions of others.\footnote{Several studies have also compared the performance of various mechanisms in eliciting subjects’ beliefs about the actions of others (e.g., see Huck and Weizsäcker, 2002; Gächter and Renner, 2010; Trautmann and van der Kuilen, 2015). However, unlike us, these papers do not directly compare the QSR and the BSR.}
Moreover, unlike both HO and HMS, we also focus on the impact that the scoring rules have on what the experimenter observes as the subjects’ updating behavior. There is an emerging literature within economics that examines how individuals process new information, and whether they deviate from the Bayesian benchmark when updating their beliefs in response to this information (e.g., see Eil and Rao, 2011; Ertac, 2011; Grossman and Owens, 2012; Möbius et al., 2014; Barron, 2016; Coutts, 2017). Hence, if the belief elicitation mechanism affects the behavior of subjects by distorting their belief reports, then it is crucial to understand how these distortions affect the way we interpret their updating behavior.

In addition, our study also differs from both HO and HMS in that the incentives to distort belief reports are likely to be less salient in our setup for a number of reasons. First, the belief elicitation task is embedded within a larger set of tasks that subjects have to perform in our experiment, whereas in both HO and HMS this is the main task of their study. Hence, subjects in our experiment need to divide their attention across various domains where decisions are being made (i.e., both investment decisions as leaders and their beliefs as group members). Next, the implementation of the elicitation mechanisms is also different between HMS and our study. In our experiment, subjects simply enter an integer on the computer as their belief report. In HMS, subjects are given real-time feedback about the implication that their decisions have on their payoffs as they adjust their belief reports using an onscreen slider. However, such feedback is absent in our design. Moreover, unlike us, HO adopts a within-subject treatment design where subjects are exposed to both scoring rules and are more likely to be able to directly compare the two. We further discuss in detail how these factors may affect the subjects’ behavior in Section 3.6.

Finally, in a different study, Harrison et al. (2015) also compare the QSR and BSR. Similar to us, they do not find evidence of distortions in belief reports under the QSR. However, they differ from our study in that they consider an environment where the scoring rules are used to elicit a distribution of subjective probabilities over continuous events, as opposed to probabilities over binary events. As suggested by Harrison et al.
(2017) in a more recent paper, the lack of differences in elicited beliefs between the QSR and the BSR may be driven by the fact that the distortions in reports under the QSR are not as severe when eliciting beliefs over continuous events as compared to those over binary events.

3.3 Experimental Design

3.3.1 Treatments

The experiment extends the design from Chapter 2, which is described in detail in Section 2.3. The main task in the experiment is the leadership task. Subjects are divided into groups of three and a leader makes an investment decision (effort choice) on behalf of the group. The leader has a choice between Investment X and Investment Y, which correspond to a high effort \(e_H\) and a low effort \(e_L\) choice, respectively. The investment can lead to either a high payoff \(Q_H\) or a low payoff \(Q_L\) for the group. The members of the group do not observe the leader’s decision but the outcome of the leader’s investment is observable. The leadership task is repeated for six rounds.

The experiment features both a within-subject and between-subject treatment design. Within subjects, the treatment variable is the method of leadership appointment (i.e., treatments RA, LA, HA, and GA, as described in Section 2.3.2), which varies across rounds. Between subjects, the treatment variable is the scoring rule used to incentivize the group members’ beliefs about the leader’s investment decision. The belief elicitation mechanism is the main treatment variable of interest in this paper.

In treatment QSR, we use the QSR to incentivize members’ beliefs about the leader’s investment decisions. Each member reports an integer \(r \in [0, 100]\), where a higher number corresponds to a higher subjective probability that the leader has chosen Investment X. The group member can receive up to 10 Experimental Currency Units (ECU) based on his belief report and the leader’s effort choice. Specifically, the group member’s payoff \(\Pi_{QSR}\) is given by

\[
\Pi_{QSR} = 10 \times \left\{ 1 - \left[ I(e = e_H) - \frac{r}{100} \right]^2 \right\}, \tag{3.1}
\]
where $I(e = e_H)$ equals 1 if the leader has chosen high effort (i.e., Investment X) and 0 otherwise. To illustrate, suppose the subject reports $r = 70$. Then, his payoff under the QSR would be 9.1 ECU if the leader chose $e_H$ and 5.1 ECU if the leader chose $e_L$. Subjects are told that, under this payment mechanism, they can expect to receive the highest payoff if they report their true beliefs about the leader’s decision.

In treatment BSR, group members’ beliefs are incentivized using the BSR. Under this scoring rule, the member receives a fixed reward (10 ECU) with probability $\pi_{BSR}$ given by

$$
\pi_{BSR} = 1 - \left[ I(e = e_H) - \frac{r}{100} \right]^2.
$$

To illustrate, suppose the subject reports $r = 70$. Then, the member receives 10 ECU with probabilities 0.91 and 0.51 if the leader chose $e_H$ and $e_L$, respectively.\(^7\) Under this scoring rule, subjects are told that it is in their best interest to report their true beliefs about the leader’s decision.\(^8\)

### 3.3.2 Procedures

The experiments were conducted in the Experimental Economics Laboratory at the University of Melbourne ($E^2MU$) and programmed using z-tree (Fischbacher, 2007). Subjects were recruited using ORSEE (Greiner, 2015). The data collected in Chapter 2 is used for the analysis as treatment BSR in this study. We ran 9 additional sessions for treatment QSR, with 24 to 30 subjects in each session. Across both treatments, a total of 528 subjects participated in the experiment.\(^9\)

Each session lasted between 90 and 120 minutes. At the end of the experiment, sub-

\(^{7}\)Note that the parameters are chosen such that the expected payoff to the subject for any given report $r$ is the same under both scoring rules.

\(^{8}\)We provide the full set of experimental instructions for treatment BSR in Appendix A.1 of the previous chapter. The detailed instructions for Stage 3 of the leadership task, which differ between treatments QSR and BSR, can be found in Appendix B.1.

\(^{9}\)Given that we restricted our subject pool to Australian citizens in Chapter 2, we also imposed the same restriction on our subject pool in treatment QSR. However, to ensure consistency of our subject pool, we decided to drop 18 subjects from the analysis (8 from treatment QSR and 10 from treatment BSR). These subjects either had prior experience with the experiment, or had misreported their citizenship on the recruitment system and indicated in the questionnaire that they have lived in Australia for less than two years. In total, data from 510 subjects (238 from treatment QSR and 272 from treatment BSR) are used for the analysis in this paper.
jects were invited to complete a brief questionnaire which included demographic questions, questions about their decisions during the experiment, and an incentivized one-shot risk game (Gneezy and Potters, 1997) to elicit their risk preferences. Subjects were paid for either the dictator game or the leadership task. Earnings were converted to cash at the conclusion of the session at the rate 10 ECU = 1 AUD. Overall, in the additional sessions for treatment QSR, subjects earned between $10 and $70, with the mean earnings being $32.97. Subjects’ earnings also included a show-up fee of $10.

3.4 Theory and Hypotheses

3.4.1 Interim beliefs

In this section, we provide a simple theoretical framework that allows us to form predictions for the behavior of group members under each of the two scoring rules.

Assume that individuals are subjective expected utility maximizers with utility given by \( u(\cdot) \), where \( u'(\cdot) > 0 \) and \( u''(\cdot) \leq 0 \). Let \( p \) denote the group member’s true subjective probability that an event occur. In our experiment, this is their belief that the leader has chosen high effort \( (e_H) \) as opposed to low effort \( (e_L) \). The member states a report \( r \in [0, 1] \) given his belief \( p \) and the scoring rule. If a scoring rule incentivizes truth-telling, then this implies that it is an optimal strategy for the member to report \( r = p \) under the scoring rule.

**Reported beliefs under the QSR.** Under the QSR, the group member receives \( \Pi_{QSR} = V \{1 - [I(e = e_H) - r]^2\} \) given his belief report \( r \) and the leader’s effort choice \( e \in \{e_L, e_H\} \), with \( V > 0 \). Normalizing \( V = 1 \), the subject solves\(^{10}\)

\[
\max_r p \ u(1 - (1 - r)^2) + (1 - p) \ u(1 - r^2),
\]

\(^{10}\)Note that this specification assumes that the utility members receive from the belief-elicitation task is additive separable from the utility they receive from the outcome of the investment chosen by the leader in the experiment. Members do not have direct control over the decisions made by the leader. Hence, the utility that they receive from the leader’s investment decision is not included in the specification.
which gives the following first-order condition (FOC)

\[ p u'(1 - (1 - r)^2)(1 - r) = (1 - p) u'(1 - r^2)r. \]  

(3.4)

For a risk-neutral member, \( u''(\cdot) = 0 \), implying that \( u'(\cdot) \) is the same on both sides of (3.4). Hence, the FOC reduces to \( p(1 - r) = (1 - p)r \), giving \( r = p \). A risk-neutral member truthfully reports his belief under the QSR.

For a risk-averse member, \( u''(\cdot) < 0 \). Note that \( 1 - (1 - r)^2 \) is increasing in \( r \). Hence, \( u'(1 - (1 - r)^2) \) on the left-hand side (LHS) of (3.4) is decreasing in \( r \) since \( u''(\cdot) < 0 \). As \( 1 - r^2 \) is also decreasing in \( r \), the entire LHS of (3.4) is decreasing in \( r \). Consequently, the entire right-hand side (RHS) of (3.4) is increasing in \( r \).

To evaluate the optimal strategy for a risk-averse member under the QSR, we first assume that \( r = p \) and check if the FOC holds. Note that this premise implies \( p(1 - r) = (1 - p)r \), and therefore the equality in (3.4) depends on the terms in \( u'(\cdot) \) on both sides of the FOC. We consider three cases. First, suppose \( r = p = \frac{1}{2} \). Then, \( 1 - (1 - r)^2 = 1 - r^2 = \frac{3}{4} \) and the equality in (3.4) holds. The member truthfully reports his belief if he thinks that the leader is equally likely to have chosen \( e_H \) and \( e_L \).

Next, suppose \( p > \frac{1}{2} \) while still maintaining the assumption that \( r = p \). Hence, the member believes that the leader is more likely to have chosen \( e_H \) than \( e_L \). Note that \( 1 - (1 - r)^2 > 1 - r^2 \) for \( r \in \left( \frac{1}{2}, 1 \right] \). Hence, \( u'(1 - (1 - r)^2) < u'(1 - r^2) \) since \( u''(\cdot) < 0 \). Consequently, the LHS of (3.4) is less than the RHS when \( p = r > \frac{1}{2} \). Since the LHS is decreasing in \( r \) and the RHS is increasing in \( r \), \( r \) will need to decrease for the FOC to hold. As a result, \( r \) decreases towards \( \frac{1}{2} \), and the FOC holds for some \( r \in \left( \frac{1}{2}, p \right) \).

By a similar argument, now suppose \( p < \frac{1}{2} \) while still maintaining the assumption that \( r = p \). Hence, the member believes that the leader is less likely to have chosen \( e_H \) than \( e_L \). Note that \( 1 - (1 - r)^2 < 1 - r^2 \) for \( r \in [0, \frac{1}{2}) \). Hence, \( u'(1 - (1 - r)^2) > u'(1 - r^2) \). Consequently, the LHS of (3.4) is now greater than the RHS when \( p = r < \frac{1}{2} \). This implies that the LHS of (3.4) is greater than the RHS when \( r = \frac{1}{2} \), and therefore the inequality will now be in the opposite direction.

\[ ^{11} \text{Note that } r \text{ will not be distorted to a value below } \frac{1}{2} \text{ because both } u'(1 - (1 - r)^2) > u'(1 - r^2) \text{ and } p(1 - r) > (1 - p)r \text{ when } r = \frac{1}{2} \text{ and } p > \frac{1}{2}. \]
implies that $r$ will need to increase for the FOC to hold. As a result, $r$ increases towards \( \frac{1}{2} \), and the FOC holds for some $r \in \langle p, \frac{1}{2} \rangle$.

Overall, under the QSR, a risk-neutral member will report $r = p$. However, a risk-averse group member will only do so if $p = \frac{1}{2}$. A risk-averse group member who holds a belief $p \neq \frac{1}{2}$ will have an incentive to distort $r$ away from $p$ and towards $\frac{1}{2}$. Intuitively, this distortion occurs because the QSR penalizes the member more on the margin for an incorrect prediction of the leader’s decision as his belief report becomes more extreme (i.e., when $r$ gets closer to either 0 or 1). Consequently, subjects who are risk averse would prefer to smooth out their payoffs by distorting their reports towards the midpoint.

**Reported beliefs under the BSR.** Under the BSR, the group member receives $V > 0$ with probability $\pi_{BSR} = 1 - [I(e = c_H) - r]^2$ given his report $r$ and the leader’s effort choice $e$. The subject solves

$$\max_r p[1 - (1 - r)^2]u(V) + (1 - p)(1 - r^2) u(V),$$

which implies $r = p$. Hence, under the BSR, a subjective expected utility maximizer will truthfully report his belief independent of his risk preferences.

**Predictions for reported interim beliefs.** In our experiment, each member reports his belief about the leader’s effort choice as an integer on the interval $[0, 100]$. From the discussion above, we expect group members to truthfully report their beliefs under the BSR independent of their risk preferences. However, under the QSR, if members are risk averse, then they will have an incentive to distort their belief reports away from their true beliefs and towards 50. Evidence from the literature suggests that subjects tend to exhibit risk-averse behavior in the laboratory (e.g., Gneezy and Potters, 1997; Holt and Laury, 2002; Eckel and Grossman, 2002, 2008). Hence, we expect subjects to distort
their belief reports in treatment QSR, and that beliefs elicited under the QSR will be closer to 50 on average than those elicited under the BSR. We summarize our prediction as follows.

**Hypothesis 3.1:** The group members’ interim beliefs of the leader’s decision reported in treatment QSR will be, on average, closer to 50 than those reported in treatment BSR.

### 3.4.2 Updating behavior

From Chapter 2, we find that group members significantly deviate from Bayes’ rule and suffer from biases in their updating behavior. In this section, we use the data obtained under the BSR in Chapter 2 and evaluate what the biases would have been if those beliefs were elicited under the QSR instead.

Consider the econometric specification presented in Section 2.5 of Chapter 2. Denote $\mu^\Psi_i$ as the group member’s interim belief of the likelihood that the leader has chosen $e_H$, given an appointment mechanism $\Psi$. $\phi^\Psi_i|Q$ is the (non-Bayesian) member’s posterior belief conditional on observing $Q \in \{Q_L, Q_H\}$. Given a high effort choice, $p$ is the probability that the investment yields $Q_H$. We estimate the augmented Bayes’ rule given by

$$\logit(\hat{\phi}^\Psi_i|Q) = \delta \logit(\mu^\Psi_i) + \gamma_G I(Q = Q_H) \cdot \logit(p) + \gamma_B I(Q = Q_L) \cdot \logit(1 - p),$$  

(3.7)

where $I(\cdot)$ is an indicator function, and the parameter estimates of $\delta$, $\gamma_G$, and $\gamma_B$ capture deviations from Bayes’ rule.$^{12}$

When beliefs are incentivized using the QSR, group members who are risk averse will distort their reports of both the interim and posterior beliefs. As a result, $\hat{\phi}^\Psi_i$ and $\mu^\Psi_i$ on both sides of (3.7) as observed by the experimenter under the QSR do not necessarily reflect the true subjective probabilities held by the members. Hence, the question is whether, and how, distortions in the reports of both interim and posterior beliefs under the QSR will bias the parameter estimates in (3.7).

To answer this question, we start with the premise that subjects truthfully report

---

$^{12}$The interpretation of the parameters are explained in detail in Section 2.5 of the previous chapter.
both their interim and posterior beliefs under the BSR, and examine what happens when they distort their belief reports under the QSR. This implies that the estimates of $\delta$, $\gamma_G$, and $\gamma_B$ that we obtained in Chapter 2 reflect the actual biases that group members suffer from in their belief-updating process.

Next, we assume that subjects have a CRRA utility function which takes the form

$$u(c) = \begin{cases} 
\frac{c^{1-\eta}}{1-\eta} & \text{if } \eta \neq 1 \\
\log c & \text{if } \eta = 1
\end{cases},$$

(3.8)

where $\eta$ is the risk-aversion coefficient. $\eta = 0$ corresponds to a risk-neutral individual while $\eta > 0$ corresponds to an individual who is risk averse. Given that $u'(c) = c^{-\eta}$, (3.4) reduces to

$$\frac{p}{1-p} = \frac{r}{1-r} \left[ \frac{1 - (1 - r)^2}{1 - r^2} \right]^\eta.$$  

(3.9)

To obtain an estimate of $\eta$, we use the subjects’ behavior in the one-shot risk game implemented in the post-experimental questionnaire. In the risk game, subjects are given an endowment of $\omega = 50$ and they can invest any portion of this amount $x$ in an asset which pays $3x$ with probability $\frac{1}{2}$ and zero otherwise. The subject keeps any portion of the endowment that is not invested. Hence, given the CRRA utility function, the subject’s risk-aversion coefficient is given by

$$\eta = \frac{\log 2}{\log \left( \frac{\omega + 2x}{\omega - x} \right)}.$$  

(3.10)

Using both (3.9) and (3.10), we calculate what the group members in treatment BSR

\footnote{An alternative to doing this is to randomly draw a risk-aversion coefficient for each individual from a given distribution. However, from the data obtained in Chapter 2, we find that subjects who are more risk averse also tend to contribute more in the dictator game, and their dictator game behavior is also a predictor of their interim beliefs. As a result, risk-averse subjects tend to have higher interim beliefs. Given this, the extent of the distortions in belief reports under the QSR is likely to depend on the distribution of the true interim beliefs held by the subjects. That is, we would expect more distortions to occur across the range of high interim beliefs since subjects holding these beliefs are also more likely to be risk averse. To address this potential confound, we decided to use the subjects’ risk-aversion coefficient as implied from their behavior in the risk game. This allows us to obtain a more reliable estimate of what the distortions in belief reports would have been under the QSR.}

\footnote{Note that $\eta$ is not defined for $x = 0$ or $x = \omega$. Hence, we let $x = 0.01$ and $x = 0.99$ for subjects choosing to invest none ($x = 0$) and all ($x = \omega$) of their endowment, respectively.}
would have reported as their interim and posterior beliefs under the QSR instead. Then, we estimate (3.7) using ordinary least squares (OLS) with these simulated belief reports, and compare the coefficient estimates to those obtained under the BSR. We do this both at the pooled level and for each appointment mechanism. The results are presented in Table 3.1.

Table 3.1: Actual updating behavior (BSR) vs. implied updating behavior (QSR)

<table>
<thead>
<tr>
<th>Parameters</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pooled</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(a) Beliefs elicited using the BSR

\[ \delta \]  
0.695 0.764 0.692 0.703 0.529

\[ \gamma_G \]  
0.751 0.744 0.622 0.847 0.798

\[ \gamma_B \]  
0.966 0.932 1.058 0.946 0.876

\[ \gamma_G - \gamma_B \]  
-0.215 -0.188 -0.437 -0.099 -0.078

(b) Implied beliefs under the QSR

\[ \delta \]  
0.714 0.787 0.718 0.691 0.559

\[ \gamma_G \]  
0.593 0.571 0.499 0.690 0.632

\[ \gamma_B \]  
0.762 0.713 0.835 0.743 0.682

\[ \gamma_G - \gamma_B \]  
-0.169 -0.142 -0.336 -0.053 -0.050

(c) Difference (QSR – BSR)

\[ \delta \]  
0.019 0.023 0.026 -0.011 0.030

\[ \gamma_G \]  
-0.157 -0.173 -0.123 -0.158 -0.166

\[ \gamma_B \]  
-0.204 -0.219 -0.224 -0.203 -0.194

\[ \gamma_G - \gamma_B \]  
0.046 0.046 0.101 0.045 0.028

Panel (a) of the table presents the coefficient estimates from the beliefs elicited under the BSR (which are the same as those reported in Table 2.4 of Chapter 2). Panel (b) presents the coefficient estimates using the belief reports that would have been obtained under the QSR. Finally, panel (c) of the table presents the differences in the estimated coefficients between panels (a) and (b). Hence, panel (c) tells us the biases we would expect to find in the coefficient estimates as a result of misreported beliefs under the QSR.

We observe from panel (c) that, at the pooled level (column 1), distortions in belief reports under the QSR lead to a slight increase in the estimates of \( \delta \), but attenuate the
estimates of $\gamma_G$ and $\gamma_B$. Moreover, the difference between $\gamma_G$ and $\gamma_B$ is smaller under the QSR than under the BSR (i.e., the difference is 0.169 under the QSR and 0.215 under the BSR). Under each appointment mechanism, the direction of the biases is similar to those at the pooled level. The one exception is that, under appointment mechanism HA, the parameter estimate for $\delta$ is lower under the QSR than under the BSR.$^{15}$

Hence, our predictions for the members’ updating behavior are as follows:

**Hypothesis 3.2:**

(i) At the pooled level, the interim and posterior beliefs elicited in treatment QSR will, relative to that in treatment BSR,

(a) increase the estimate of $\delta$;

(b) attenuate the estimates of both $\gamma_G$ and $\gamma_B$; and

(c) decrease the difference between the estimates of $\gamma_G$ and $\gamma_B$.

(ii) Under each appointment mechanism, the differences in parameter estimates between treatments QSR and BSR are similar to that described in (i). However, under appointment mechanism HA, the estimate of $\delta$ will be lower in treatment QSR than in treatment BSR.

### 3.5 Results

#### 3.5.1 Decisions in the risk game

Before we discuss the main results, we first analyze the subjects’ behavior under the risk game. This is important since the theoretical predictions discussed in Section 3.4 are driven by the subjects’ risk preferences. Hence, we first examine if there are any systematic differences in subjects’ risk preferences between the two treatments which may otherwise affect the interpretation of our results.

$^{15}$This difference may be driven by the fact that, under mechanism HA, the subjects’ interim beliefs are much higher. As discussed in footnote 13, since risk-averse subjects are more likely to have higher interim beliefs, it is possible that the distortions in belief reports under this appointment mechanism are likely to be higher as well.
Figure 3.1: Decisions in risk game (QSR vs. BSR)

Figure 3.1 presents the distribution of subjects’ investment decisions in the risk game for treatments QSR and BSR. From the figure, we can see that the subjects’ decisions in the risk game do not appear to be different between the two treatments. A Kolmogorov-Smirnov test for the equality of distribution functions fails to reject the null hypothesis that the distribution of investment decisions is different between the two treatments (p-value = 0.882). Out of an endowment of 50, the average investment is 34.92 in treatment QSR and 36.05 in treatment BSR. However, a Wilcoxon rank-sum test fails to reject the null hypothesis that the subjects’ behavior in the risk game is different between the BSR and QSR treatments (p-value = 0.418).

Overall, we conclude that there is no statistically significant evidence to suggest that subjects between the two samples differ in their risk preferences on average.

3.5.2 Interim beliefs

Overview

In this section, we compare the reported interim beliefs between treatments QSR and BSR. Figure 3.2 shows the distribution of reported interim beliefs, both at the pooled
level (panel a), and for each appointment mechanism (panels b to e). The dashed line in each panel corresponds to a belief of 50.

Panel (a) of Figure 3.2 suggests that, at the pooled level, the interim beliefs do not appear to be different between treatments QSR and BSR on average. A Kolmogorov-
Belief Elicitation: QSR vs. BSR

The Smirnov test fails to reject the null hypothesis that the distribution of elicited beliefs is different between treatments (p-value = 0.278).

Panel (c) reveals that, under appointment mechanism LA, the reported interim beliefs are skewed more to the right in treatment QSR than in treatment BSR. A Kolmogorov-Smirnov test reveals that the distribution of beliefs is marginally statistically significantly different between these two treatments (p-value = 0.093). We also observe that, for the other appointment mechanisms, there appear to be more belief reports closer to 50 in treatment QSR than in treatment BSR. Nonetheless, Kolmogorov-Smirnov tests fail to reject the null hypotheses that the distribution of beliefs is different between treatments QSR and BSR for each of these appointment mechanisms (RA: p-value = 0.296; HA: p-value = 0.354; GA: p-value = 0.983).

Overall, we do not find any statistically significant evidence to suggest that the reported interim beliefs are different between treatments QSR and BSR.\textsuperscript{16}

**Distance of interim beliefs from 50**

In this section, we conduct a direct test of Hypothesis 3.1. The key variable of interest is the absolute distance of the reported interim belief from 50. Figure 3.3 presents the empirical cumulative distributions of the distance of reports from 50, both at the pooled level (panel a), and for each appointment mechanism (panels b to e).

Panel (a) of Figure 3.3 reveals that the QSR does not lead to belief reports that are closer to 50 than the BSR on average. At the pooled level, the average distance of the reports from 50 is 24.20 and 23.70 in treatments QSR and BSR, respectively. A Wilcoxon rank-sum test fails to reject the null hypothesis that the distance of belief reports is different between these two treatments (p-value = 0.542).

Similar observations can be made for each appointment mechanism. Panels (b) to

\textsuperscript{16}This result is also robust to regression analyses of the members' interim beliefs. While we do not uncover any systematic differences in the beliefs elicited under the two scoring rules, an important question is whether there are any differences in the impact that the appointment mechanisms have on the reported interim beliefs. That is, we are interested to find out if the elicitation mechanism affects the conclusions in Result 2.1 of Chapter 2. To test this, we also include as controls in our regression analyses the leadership appointment mechanisms. We find that there are no statistically significant differences in the impact of the leadership appointment mechanism on the reported interim beliefs between treatments QSR and BSR. Details of these additional analyses can be found in Appendix B.2.
Figure 3.3: Empirical cumulative distributions of absolute distance of reported interim beliefs from 50 (QSR vs. BSR)

(e) reveal that, with the exception of mechanism LA, the distance of reported interim beliefs from 50 is no different between treatments QSR and BSR on average. Panel (c) of Figure 3.3 reveals that, under appointment mechanism LA, the distance of reports is
smaller on average in treatment BSR than in treatment QSR. Under this mechanism, the average distance of the reports from 50 is 28.88 and 26.72 in treatments QSR and BSR, respectively. A Wilcoxon rank-sum test reveals that this difference in mechanism LA is statistically significant (p-value = 0.029). However, for appointment mechanisms RA, HA, and GA, there are no statistically significant differences in the distance of reported interim beliefs between treatments QSR and BSR (Wilcoxon rank-sum tests: (i) RA: p-value = 0.429; (ii) HA: p-value = 0.397; and (iii) GA: p-value = 0.934).

Table 3.2 presents OLS estimates from the regression of the distance of belief reports from 50 against the scoring rule (i.e., QSR vs. BSR). In all specifications, treatment BSR is the reference treatment. In the regression analyses, we control for subjects’ behavior in the dictator and risk games, the appointment mechanisms in the leadership task (of which RA is the reference mechanism), and subjects’ characteristics, which include age, gender, study level at University, whether the subject is pursuing a major in economics, and previous experience with economic experiments.

Column (1) of Table 3.3 reveals that the average distance of interim beliefs is smaller in treatment QSR than in treatment BSR, although this difference is small and statistically insignificant (p-value = 0.750). This treatment effect remains statistically insignificant even after including controls for the appointment mechanisms in the leadership task (column 2, p-value = 0.750) and interacting the scoring rule with the subject’s behavior in the risk game (column 3, p-value = 0.709).\textsuperscript{17}

Overall, we fail to find support for Hypothesis 3.1. We summarize our result as follows.

\textbf{Result 3.1:} There is no statistically significant evidence to suggest that the reported interim beliefs elicited under the QSR are closer to 50 on average than those elicited under the BSR.

\textsuperscript{17}We also observe that the subject’s behavior in the risk game has an impact on the distance of belief reports. Specifically, a subject who exhibits more risk-averse behavior (by investing less in the risk game) is more likely to report a belief closer to 50 on average. However, this effect is marginally statistically significant in columns (1) and (2) (p-value = 0.093 for both specifications), and statistically insignificant once we include interaction terms with the belief elicitation mechanism in column (3) (p-value = 0.284).
Table 3.2: Regression of absolute distance of interim belief from 50

<table>
<thead>
<tr>
<th>Variables</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment QSR</td>
<td>-0.300</td>
<td>-0.300</td>
<td>-0.929</td>
</tr>
<tr>
<td></td>
<td>(0.940)</td>
<td>(0.941)</td>
<td>(2.489)</td>
</tr>
<tr>
<td>% endowment transferred in DG</td>
<td>-0.058**</td>
<td>-0.058**</td>
<td>-0.058**</td>
</tr>
<tr>
<td></td>
<td>(0.025)</td>
<td>(0.025)</td>
<td>(0.025)</td>
</tr>
<tr>
<td>% endowment invested in RG</td>
<td>0.028*</td>
<td>0.028*</td>
<td>0.024</td>
</tr>
<tr>
<td></td>
<td>(0.017)</td>
<td>(0.017)</td>
<td>(0.022)</td>
</tr>
<tr>
<td>Appointment mechanism LA</td>
<td>6.468***</td>
<td>6.468***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.747)</td>
<td>(0.747)</td>
<td></td>
</tr>
<tr>
<td>Appointment mechanism HA</td>
<td>1.695**</td>
<td>1.695**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.688)</td>
<td>(0.688)</td>
<td></td>
</tr>
<tr>
<td>Appointment mechanism GA</td>
<td>-0.284</td>
<td>-0.284</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.699)</td>
<td>(0.699)</td>
<td></td>
</tr>
<tr>
<td>Treatment QSR × % invested in RG</td>
<td>0.009</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.032)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>0.094</td>
<td>0.094</td>
<td>0.097</td>
</tr>
<tr>
<td></td>
<td>(0.080)</td>
<td>(0.080)</td>
<td>(0.081)</td>
</tr>
<tr>
<td>Female</td>
<td>-1.882*</td>
<td>-1.882*</td>
<td>-1.901*</td>
</tr>
<tr>
<td></td>
<td>(0.968)</td>
<td>(0.968)</td>
<td>(0.968)</td>
</tr>
<tr>
<td>Trans/intersex/other</td>
<td>-1.542</td>
<td>-1.542</td>
<td>-1.494</td>
</tr>
<tr>
<td></td>
<td>(5.055)</td>
<td>(5.057)</td>
<td>(5.066)</td>
</tr>
<tr>
<td>Economics major</td>
<td>-0.965</td>
<td>-0.965</td>
<td>-0.965</td>
</tr>
<tr>
<td></td>
<td>(1.512)</td>
<td>(1.513)</td>
<td>(1.515)</td>
</tr>
<tr>
<td>First year undergraduate student</td>
<td>0.988</td>
<td>0.988</td>
<td>0.995</td>
</tr>
<tr>
<td></td>
<td>(1.122)</td>
<td>(1.123)</td>
<td>(1.124)</td>
</tr>
<tr>
<td>Undergraduate student = 1</td>
<td>-0.487</td>
<td>-0.487</td>
<td>-0.394</td>
</tr>
<tr>
<td></td>
<td>(2.727)</td>
<td>(2.728)</td>
<td>(2.771)</td>
</tr>
<tr>
<td>Postgraduate student = 1</td>
<td>0.623</td>
<td>0.623</td>
<td>0.671</td>
</tr>
<tr>
<td></td>
<td>(2.551)</td>
<td>(2.552)</td>
<td>(2.578)</td>
</tr>
<tr>
<td># previous experiments</td>
<td>0.202</td>
<td>0.202</td>
<td>0.203</td>
</tr>
<tr>
<td></td>
<td>(0.219)</td>
<td>(0.219)</td>
<td>(0.219)</td>
</tr>
<tr>
<td>Constant</td>
<td>22.442***</td>
<td>19.769***</td>
<td>19.958***</td>
</tr>
<tr>
<td></td>
<td>(4.329)</td>
<td>(4.368)</td>
<td>(4.349)</td>
</tr>
<tr>
<td>Observations</td>
<td>3.060</td>
<td>3.060</td>
<td>3.060</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.023</td>
<td>0.056</td>
<td>0.056</td>
</tr>
</tbody>
</table>

Robust standard errors clustered at the subject level in parentheses. For all regressions, treatment BSR is the reference treatment. Appointment mechanism RA is the reference mechanism in columns (2) and (3).
DG: Dictator Game; RG: Risk Game.
*** p<0.01, ** p<0.05, * p<0.10.
3.5.3 Updating behavior

We now estimate equation (3.7) using OLS to compare the biases in updating behavior based on the reported beliefs in treatments QSR and BSR. Table 3.3 presents the regression results both at the pooled level (column 1) and for each appointment mechanism (columns 2 to 5). Panels (a) and (b) of the table present the results for treatments BSR and QSR, respectively. Note that panel (a) of Table 3.3 is the same as Table 2.4 in Chapter 2.18

The coefficient estimates in column (1) reveal that, at the pooled level, we observe similar biases in the members’ belief-updating process between treatments BSR and QSR. In both treatments, members suffer from base-rate neglect (i.e., $\delta < 1$: (i) BSR: p-value < 0.001; (ii) QSR: p-value < 0.001), and they attribute good outcomes more to luck than a Bayesian would (i.e., $\gamma_G < 1$: (i) BSR: p-value < 0.001; (ii) QSR: p-value = 0.003). While group members respond to bad outcomes no differently from a Bayesian in treatment BSR (i.e., $\gamma_B = 1$: p-value = 0.608), we observe that they attribute bad outcomes more to effort than a Bayesian would in treatment QSR, although this effect is marginally statistically significant (i.e., $\gamma_B > 1$: p-value = 0.090). Nonetheless, we observe that members attribute good (bad) outcomes more to luck (effort) across both treatments (i.e., $\gamma_G < \gamma_B$), and this asymmetry is statistically significant (BSR: p-value = 0.002; QSR: p-value < 0.001).

When we compare the coefficient estimates between treatments BSR and QSR in column (1), we observe that, compared to treatment BSR, the estimate of $\delta$ is lower in treatment QSR but that of both $\gamma_G$ and $\gamma_B$ are higher. This is in the opposite direction of what we predict in Hypothesis 3.2(i). The differences in estimates between treatments BSR and QSR are not statistically significant for $\delta$ (p-value = 0.137) and $\gamma_G$ (p-value = 0.878), but is marginally statistically significant for $\gamma_B$ (p-value = 0.079). Moreover, also contrary to Hypothesis 3.2(i), the difference between $\gamma_G$ and $\gamma_B$ is larger in treatment

---

18Following the analysis from Chapter 2, we exclude subjects classified as inconsistent or non-updaters. The same definition is used to classify subjects across both treatments. Specifically, subjects are classified as inconsistent if at least 25% of their belief updates are in a direction opposite to the observed outcome, and as non-updaters if they report posterior beliefs equal to their interim beliefs across all appointment mechanisms. This implies excluding 24.6% and 26.1% of the subjects from the analysis in treatments BSR and QSR, respectively.
Table 3.3: Regression of members’ posterior beliefs (BSR vs. QSR)

<table>
<thead>
<tr>
<th>Variables</th>
<th>(1) Pooled</th>
<th>(2) RA</th>
<th>(3) LA</th>
<th>(4) HA</th>
<th>(5) GA</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Treatment BSR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(\delta) : Logit(interim belief)</td>
<td>0.695***</td>
<td>0.764***</td>
<td>0.692***</td>
<td>0.703***</td>
<td>0.529***</td>
</tr>
<tr>
<td></td>
<td>(0.039)</td>
<td>(0.071)</td>
<td>(0.054)</td>
<td>(0.058)</td>
<td>(0.135)</td>
</tr>
<tr>
<td>(\gamma_G) : Good outcome (\times) (\logit(p))</td>
<td>0.751***</td>
<td>0.744***</td>
<td>0.622***</td>
<td>0.847*</td>
<td>0.798**</td>
</tr>
<tr>
<td></td>
<td>(0.051)</td>
<td>(0.089)</td>
<td>(0.079)</td>
<td>(0.081)</td>
<td>(0.098)</td>
</tr>
<tr>
<td>(\gamma_B) : Bad outcome (\times) (\logit(1-p))</td>
<td>0.966</td>
<td>0.932</td>
<td>1.058</td>
<td>0.946</td>
<td>0.876</td>
</tr>
<tr>
<td></td>
<td>(0.067)</td>
<td>(0.092)</td>
<td>(0.117)</td>
<td>(0.072)</td>
<td>(0.114)</td>
</tr>
<tr>
<td>Observations</td>
<td>2,460</td>
<td>410</td>
<td>820</td>
<td>820</td>
<td>410</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.608</td>
<td>0.686</td>
<td>0.651</td>
<td>0.583</td>
<td>0.421</td>
</tr>
</tbody>
</table>

Test of \(\gamma_G = \gamma_B\)

<table>
<thead>
<tr>
<th>t-statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>-3.190</td>
<td>0.002</td>
</tr>
</tbody>
</table>

(b) Treatment QSR

<table>
<thead>
<tr>
<th>Variables</th>
<th>(1) Pooled</th>
<th>(2) RA</th>
<th>(3) LA</th>
<th>(4) HA</th>
<th>(5) GA</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\delta) : Logit(interim belief)</td>
<td>0.595***</td>
<td>0.549***</td>
<td>0.641***</td>
<td>0.521***</td>
<td>0.587***</td>
</tr>
<tr>
<td></td>
<td>(0.055)</td>
<td>(0.103)</td>
<td>(0.061)</td>
<td>(0.086)</td>
<td>(0.085)</td>
</tr>
<tr>
<td>(\gamma_G) : Good outcome (\times) (\logit(p))</td>
<td>0.765***</td>
<td>0.807</td>
<td>0.645***</td>
<td>0.982</td>
<td>0.737**</td>
</tr>
<tr>
<td></td>
<td>(0.077)</td>
<td>(0.122)</td>
<td>(0.093)</td>
<td>(0.140)</td>
<td>(0.113)</td>
</tr>
<tr>
<td>(\gamma_B) : Bad outcome (\times) (\logit(1-p))</td>
<td>1.230*</td>
<td>1.231</td>
<td>1.074</td>
<td>1.303**</td>
<td>1.189</td>
</tr>
<tr>
<td></td>
<td>(0.135)</td>
<td>(0.177)</td>
<td>(0.165)</td>
<td>(0.141)</td>
<td>(0.159)</td>
</tr>
<tr>
<td>Observations</td>
<td>2,112</td>
<td>352</td>
<td>704</td>
<td>704</td>
<td>352</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.500</td>
<td>0.414</td>
<td>0.629</td>
<td>0.389</td>
<td>0.484</td>
</tr>
</tbody>
</table>

Test of \(\gamma_G = \gamma_B\)

<table>
<thead>
<tr>
<th>t-statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>-4.216</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Robust standard errors clustered at the subject level in parentheses. This analysis excludes subjects classified as inconsistent or non-updaters.

*** p<0.01, ** p<0.05, * p<0.10. Null hypothesis is coefficient = 1.
BELIEF ELICITATION: QSR VS. BSR

QSR than in treatment BSR, although this effect is marginally statistically significant (p-value = 0.053). Overall, at the pooled level, we fail to find systematic support for Hypothesis 3.2(i). We summarize our result as follows.

**Result 3.2:** There are no statistically significant differences in the estimates of $\delta$ and $\gamma_G$ between treatments QSR and BSR at the pooled level. However, contrary to Hypothesis 3.2(i), the estimate of $\gamma_B$, as well as the difference between that of $\gamma_G$ and $\gamma_B$, are both marginally statistically significantly higher under the QSR than under the BSR.

We now examine the coefficient estimates for each appointment mechanism. In treatment BSR, we observe similar biases in the updating behavior across all four appointment mechanisms. Members consistently suffer from base-rate neglect ($\delta < 1$), attribute good outcomes more to luck than a Bayesian would ($\gamma_G < 1$), but treat bad outcomes no differently from a Bayesian ($\gamma_B = 1$).

In treatment QSR, we also observe that group members consistently suffer from base-rate neglect under each of the four appointment mechanism ($\delta < 1$: p-values < 0.001 for all appointment mechanisms). However, the attribution of outcomes depends on the appointment mechanism in treatment QSR. Specifically, members attribute good outcomes more to luck than a Bayesian would under mechanisms LA and GA (p-value < 0.001 and p-value = 0.021, respectively), but are no different from a Bayesian under mechanisms RA and HA (p-values = 0.116 and 0.896, respectively). Moreover, while members treat bad outcomes no differently from a Bayesian in mechanisms RA, LA, and GA (p-values = 0.193, 0.654, and 0.238, respectively), they attribute bad outcomes more to effort than a Bayesian would under mechanism HA (p-values = 0.034).

In addition, we observe that, in treatment QSR, members consistently attribute good (bad) outcomes more to luck (effort) under each appointment mechanism ($\gamma_G < \gamma_B$). This effect is statistically significant across all four mechanisms (RA: p-value = 0.019; LA: p-value = 0.038; HA: p-value = 0.017; GA: p-value = 0.007). However, this effect is only statistically significant under mechanism LA in treatment BSR (RA: p-value = 0.114; LA: p-value = 0.002; HA: p-value = 0.281; GA: p-value = 0.609).

As a direct test of Hypothesis 3.2(ii), pairwise comparisons of the coefficient estimates
between treatments BSR and QSR for each appointment mechanism reveal the following statistically significant differences. First, the estimates for $\delta$ are lower in treatment QSR than in treatment BSR under mechanisms RA and HA. While these differences are in the same direction as predicted in Hypothesis 3.2(ii), they are only marginally statistically significant (RA: p-value = 0.086; HA: p-value = 0.080).

Next, the estimate for $\gamma_B$ is statistically significantly higher in treatment QSR than in treatment BSR under mechanism HA (p-value = 0.025). Moreover, the difference between the estimates of $\gamma_G$ and $\gamma_B$ is marginally statistically significantly higher in treatment QSR than in treatment BSR under mechanism GA (p-value = 0.098). However, we note that both of these differences are in the opposite direction to our predictions in Hypothesis 3.2(ii).

Overall, we fail to find systematic support for Hypothesis 3.2(ii).

**Result 3.3:** There is no statistically significant evidence to suggest that there are systematic differences in updating behavior between treatments QSR and BSR under each appointment mechanism.

### 3.6 Discussion

This paper evaluates the performance of the quadratic and binarized scoring rules in eliciting subjects’ beliefs about the decisions made by other individuals. Contrary to our theoretical predictions, we do not find any statistically significant differences in the interim beliefs between the two scoring rules. Moreover, we do not observe any systematic differences in the biases that members suffer from between the two treatments. Our findings are also different from those of Hossain and Okui (2013) and Harrison et al. (2014). In this section, we discuss possible reasons for why our results are different to our theoretical predictions.

**Stakes.** It is possible that the stakes used in the belief-elicitation task can affect how subjects respond to the different scoring rules between the two treatments. In our experiment, subjects are paid relatively small stakes (i.e., up to 1 AUD) for their belief reports.
On the other hand, subjects are paid up to 800 JPY (9.36 AUD) for their belief reports in Hossain and Okui (2013), while in Harrison et al. (2014) they are paid up to 50 USD (66.37 AUD). Hence, one question is whether our results could be driven by the small stakes used in our experiment.

In particular, it is possible that subjects exhibit risk-neutral behavior over small stakes but are risk averse over larger stakes (e.g., see Bombardini and Trebbi, 2012). Such behavior would be consistent with subjective expected utility maximization if the subjects take into account the stakes in the belief-elicitation task relative to the total payment they expect to receive from the experiment. If this is the case, then it is possible that subjects exhibit close to risk-neutral behavior even under the QSR in our experiment. This may in turn explain why we fail to find any differences in the belief reports between the two treatments.

To test this conjecture, one possible extension to our study is to examine whether the behavior between the QSR and BSR treatments in our environment will be different when the stakes are increased. If we observe differences in the reported beliefs between the two treatments when the payments in the belief-elicitation task are higher, then this could suggest that the lack of differences we find in this paper may be due to subjects exhibiting close to risk-neutral behavior under the QSR given the low stakes.

**Implementation and presentation of the scoring rules.** In practice, there are many different ways to present the elicitation mechanisms to the subjects. In our study, we present each scoring rule as a mathematical formula in the printed instructions, which is a method often used in the literature (e.g., see Schlag et al., 2015). Moreover, the decision screens are exactly the same between the QSR and BSR treatments, and our subjects are not actively reminded of the scoring rule throughout the entire experiment. On the other hand, unlike us, Hossain and Okui (2013) adopt a within-subject treatment design where their subjects are exposed to both the BSR and the QSR, and therefore may

---

19 For example, subjects in our experiment are guaranteed a minimum show-up fee of 10 AUD. Prior to participating in the experiment, they are informed that the average payment would be between 20 and 60 AUD for a 2-hour experiment. As such, the stakes used in the belief-elicitation task correspond to less than 5% of the total payment subjects would have expected to receive from the entire experiment.

20 The screenshots can be found in Appendix A.2 of Chapter 2.
find it easier to compare the two. In Harrison et al. (2014), their subjects can observe in real time how their payoff (QSR) or the probability of receiving the fixed payment (BSR) changes as they adjust their belief reports.

Given this, one concern is that the subjects in our study may have failed to fully understand the implication of the scoring rules on their own payoffs. That is, it is possible that they do not know what their optimization problem is and/or the optimal strategies under each scoring rule. To further examine this conjecture, we analyze the subjects’ performance in the set of control questions that they attempted prior to participating in the leadership task. We find evidence that more subjects in treatment QSR may have failed to understand the scoring rule as compared to those in treatment BSR. In particular, for the question relating to the belief-elicitation task, 27.7% of subjects answered the question incorrectly on the first attempt in treatment QSR, as compared to 12.9% in treatment BSR. This difference is statistically significant (Fisher’s exact test: p-value < 0.001). Moreover, subjects take a longer time on average to answer all the practice questions correctly in treatment QSR (5.49 minutes) than in treatment BSR (5.07 minutes), and this difference is also statistically significant (Wilcoxon rank-sum test: p-value = 0.045).

If subjects in the QSR treatment have misunderstood the implications of the scoring rule, then there are several possibilities as to how these subjects could have behaved. The first is that subjects could have simply provided random belief reports throughout the experiment. However, given that we find systematic patterns in both the interim and posterior beliefs (i.e., we find that the appointment mechanisms have an impact on the interim beliefs, and members respond to outcomes asymmetrically), this is unlikely to be the case. Another possibility is that subjects may have chosen to simply reveal their true beliefs, because either that seems to be a natural default option, or that the experimenter has asked them to do so (e.g., experimenter demand effect, see Zizzo, 2010).21 If this is the case, then it may explain why beliefs elicited under the QSR are similar to those elicited under the BSR.

21Note that this may also apply to subjects in the BSR treatment who have failed to understand the scoring rule.
In order to explore this issue further, a possible extension to our study is to consider different implementations of the elicitation mechanisms, where the impact of the scoring rules on the subjects’ payoffs is made more salient to them. One example is to provide an on-screen calculator that the subjects can use to calculate either their payoffs (under the QSR) or the probabilities of receiving the fixed payment (under the BSR) as they adjust their belief reports. This would allow us to test if an alternative implementation would increase the subjects’ understanding of the scoring rules. Importantly, it would enable us to examine whether there are greater distortions in belief reports under the QSR as compared to the BSR when subjects have a better understanding of the respective scoring rules.

Risk preferences. An important feature of our design is that we compare the performance of the QSR and the BSR in an environment where individuals are asked to evaluate the decisions made by others. In Hossain and Okui (2013) and Harrison et al. (2014), subjects are asked to state their beliefs about the composition of either an urn or a bingo cage, and strategic concerns are unlikely to be present in their setup. Due to the nature of our design, one challenge that arises is that both social and risk preferences may play a role in shaping the subjects’ beliefs about the decisions of others. Consequently, this may add a layer of complexity to our comparison of the subjects’ behavior between the two treatments.

In particular, we find that subjects’ risk preferences also affect their underlying interim beliefs. We base this observation on data from treatment BSR, where, in theory, we expect subjects to report their beliefs truthfully. From columns (1) and (3) of Table 2.3 in Chapter 2, we observe that subjects who are less risk averse (i.e., by investing more in the risk game) also tend to have lower interim beliefs on average. In addition to this, we also make the following observations: (i) Subjects’ beliefs are shaped by their own behavior as leaders (i.e., a false consensus effect). (ii) Subjects’ own behavior as leaders

Note that in Harrison et al. (2014), subjects state their belief reports by adjusting sliders on the screen. We chose not to adopt their interface in our experiment because we were worried that the slider task may be too tedious to the subjects, and may therefore distract them too much from the main task at hand.
is determined by their social preferences. (iii) Subjects’ risk and social preferences may be correlated.

We find evidence of (i) in Table 2.3 of Chapter 2, where in columns (3) and (4) we observe that subjects who exert high effort as leaders are also more likely to, as group members, believe their leader has chosen high effort. Next, (ii) can be observed from Table 2.2 of Chapter 2. The coefficient estimates in the table suggest that subjects who are more pro-social (by contributing more in the dictator game) are also more likely to choose high effort as leader. This is true both at the pooled level and under each appointment mechanism. Finally, we conclude (iii) from the Spearman’s rank correlation coefficient for subjects’ investment in the risk game and their contributions in the dictator game. This correlation coefficient is given by -0.235, which suggests that individuals who invest less in the risk game (i.e., more risk averse) also tend to contribute more in the dictator game on average. We find this correlation to be statistically significant (p-value < 0.001).

These observations jointly tell us that our environment may potentially introduce an additional confound — that there seems to be an interaction between the subjects’ risk and social preferences, which may in turn influence their behavior as leaders, and, consequently, their underlying beliefs about the behavior of other individuals. This becomes an issue when we compare the subjects’ behavior between the QSR and BSR treatments, since the theoretical predictions in Section 3.4 are driven by the subjects’ risk preferences. In other words, how social and risk preferences affect the true subjective probabilities of the subjects is likely to affect the differences we would expect to observe between the QSR and BSR treatments.

To further explore these issues, as future work we plan to consider a theoretical framework that accounts for such an interaction. For example, we could introduce social preferences in the objective functions under each scoring rule, i.e., in equations (3.3) and (3.5). This will involve making some assumptions on how these preferences may influence the members’ subjective probabilities about the leader’s decision. This can then allow us to form predictions about the behavior between the QSR and BSR treatments that
account for the interaction between the subjects’ social and risk preferences.

Additionally, we could remove or minimize the role that other-regarding preferences play in our setup. One way of doing this is to consider an environment where the investment decisions are determined randomly by a computer, and elicit subjects’ beliefs about the decisions made by the computer. This will allow us to compare the impact of the scoring rule on the subjects’ behavior without having to worry about potential confounds arising from the interaction between their risk and social preferences. Importantly, it may also allow us to understand why our findings are different to that of both Hossain and Okui (2013) and Harrison et al. (2014), and whether this difference is due to the role that social preference play in our decision-making environment.

To conclude, the results from our study do not necessarily suggest that the two scoring rules do not have a different impact on the subjects’ behavior. The response to different scoring rules or belief elicitation mechanisms may depend on other factors, such as the stakes associated with the payment mechanisms, how the mechanisms are implemented or presented to the subjects, as well as the decision-making environment in which the beliefs are elicited. More work needs to be done to understand why our results differ from the theoretical predictions and from the existing studies in the literature.
Chapter 4

Monetary and non-monetary incentives in real-effort tournaments

4.1 Introduction

The impact of incentives on the supply of effort is a central question in economics. In recent years, economists have been increasingly relying on laboratory experiments, especially experiments with real-effort tasks, to study this question. Real-effort tasks require experimental participants to exert actual physical or mental effort to earn money.\(^1\) They are seen as a better representation of how people earn their income in the real world in comparison to the alternative ‘stated effort’ designs, where subjects are given information on the costs and benefits of different effort options and are asked to state the level of effort they would like to exert (see, e.g., Charness and Kuhn, 2011).

While the use of real-effort tasks may contribute to the external validity of experimental results, it may also come at a cost of loss of control over incentives.\(^2\) The main concern voiced in the literature is that in real-effort tasks, researchers cannot observe the cost of effort which differs across subjects. This makes it challenging to compare

\(^1\)Examples include adjusting sliders on the screen (Gill and Prowse, 2012), pressing keys (Berger and Pope, 2011), counting zeroes (Abeler et al., 2011), typing paragraphs (Dickinson, 1999), cracking walnuts (Fahr and Irlenbusch, 2000), stuffing letters into envelopes (Konow, 2000; Falk and Ichino, 2006; Carpenter et al., 2010b), encrypting words (Erkal et al., 2011), adding numbers (Niederle and Vesterlund, 2007), solving mazes (Gneezy et al., 2003), picking virtual apples in a basket (Eriksson et al., 2017), and solving crossword puzzles (Kraut et al., 2005).

\(^2\)This is concerning since the key advantage of using experimental methods is the ability to control for extraneous confounding influences and, hence, the ability to make direct inferences relating to causality.
the behavior of subjects against theoretical benchmarks derived from assumed cost-of-effort functions.\textsuperscript{3} In contrast, in stated-effort experiments, the parameters guiding the optimization problem of subjects are clear.

Not knowing the (monetary) marginal cost of effort is not the only reason for loss of control in real-effort tasks. The introduction of a real-effort task also means that non-monetary incentives (such as task enjoyment) can play a larger role in the determination of effort. Hence, if the goal of the researcher is to measure the impact of monetary incentives on effort provision, clearly this becomes more challenging. The presence of non-monetary incentives may make it difficult to identify the impact of monetary incentives. This may be the reason why results from experiments using real-effort tasks provide mixed evidence on the relationship between monetary incentives and effort provision. For example, in their survey article, Camerer and Hogarth (1999) report that in some real-effort studies subjects respond to monetary incentives, while in others incentives fail to provoke a response or may even backfire (see also Jenkins et al., 1998; Gneezy and Rustichini, 2000; Pokorny, 2008; Ariely et al., 2009; Araujo et al., 2016; DellaVigna and Pope, forthcoming).

This evidence suggests that it is important to have a systematic exploration of the impact of monetary and non-monetary incentives in order to determine under what circumstances they increase effort in real-effort tasks. Our interest in studying effort provision using real-effort tasks is therefore motivated by the fact that real-effort tasks are becoming increasingly popular in the experimental literature, but there is mixed evidence of the impact of monetary incentives on real-effort provision. The goal of this paper is to identify the impact of monetary incentives on effort provision by reducing the impact of non-monetary incentives. We report findings from three related experiments designed to study effort provision within a real-effort contest setup. Contests are used in many economic, political, and social environments, e.g., competition for bonuses and promotions within firms, patent races, lobbying for government contracts, elections, political conflicts, and sports. Our approach is based on the conjecture that, in addition to money,
the following factors may contribute to participants’ effort choices in real-effort contests: (i) task enjoyment; (ii) the desire to do what is expected of one and/or something productive/useful for others (which results in the experimenter demand effect); and (iii) competitiveness.

Experiment 1 presents the ‘puzzle’ we aim to address in this paper. Subjects participate in two-player tournaments with two prizes. We eliminate the impact of monetary incentives on behavior by setting the prizes equal to each other in the baseline treatment. We show that effort provision in this treatment is not significantly different from effort provision in two other treatments with low and high monetary incentives. Hence, non-monetary incentives by themselves are sufficient to motivate subjects to exert significant levels of effort in real-effort tasks. Moreover, providing monetary incentives does not have an impact on effort provision. This result is in line with other findings in the literature which document that behavior in real-effort tasks is relatively insensitive to changes in monetary incentives (e.g., Camerer and Hogarth, 1999; Araujo et al., 2016).

In Experiment 2, we design treatments to reduce the impact that the non-monetary incentives mentioned above have on effort. Using a simple framework, we explain that this can be done either directly by reducing the benefit subjects receive from task enjoyment, etc. (e.g., by picking a task that is likely to be less enjoyable), or indirectly by increasing the cost of working on the task. We consider two treatments aimed at increasing the opportunity cost of effort (by allowing participants to leave early and by providing them with a small payment for taking a timeout), and a third treatment with an incentivized alternative activity which aims to increase the opportunity cost of effort at the same time it decreases the non-monetary benefits subjects may receive from performing the real-effort task. We show that participation and effort in the real-effort task are significantly lower in all three treatments as compared to the baseline treatment from Experiment 1. Moreover, introducing an incentivized alternative activity that pays a piece rate is the most effective way of reducing the impact of non-monetary incentives, while allowing subjects to leave early is the least effective.

In Experiment 3, we again examine the response of agents to changes in monetary
incentives. However, this time we use the design with an incentivized alternative activity from Experiment 2. We find that, in contrast to our results from Experiment 1, participants increase effort monotonically in response to monetary incentives. Hence, our results reveal that, with appropriate modifications to the way real-effort tasks are implemented in the laboratory, it is possible for the experimenter to observe changes in effort provision as the level of monetary incentives is varied.

In general, the findings from our three experiments highlight that understanding the role of incentives in real-effort tasks has important implications. We provide a framework that allows the researcher to mitigate the impact of non-monetary incentives in real-effort experiments. Beyond studying the role of monetary incentives on effort, doing this can be important if the researcher wants to examine how the process of earning one’s income affects the decisions in later stages of the experiment. If non-monetary incentives (e.g., experimenter demand effect or task curiosity) play a bigger role in determining effort choices, then subjects’ behavior in the later stage may be driven by their response to these incentives rather than differences in income levels. In this case, failing to acknowledge the role that non-monetary incentives play on effort provision in the first stage may result in a misidentification of the causal effect.

The rest of the paper proceeds as follows. In the next section, we review the related literature and discuss our contributions to it. In Sections 4.3 to 4.5, we present the design and results of Experiments 1, 2, and 3, respectively. We conclude in Section 4.6 with a discussion of the implications of our findings.

4.2 Related Literature

We build on and contribute to different strands of the literature. Our first contribution is to the literature on the impact of incentives. We examine how experimental economists can study the role of incentives using real-effort experiments in the laboratory. A central principle of economic analysis is that people respond to incentives. Hence, economists see incentives as a powerful tool for explaining and altering individual behavior. When it
comes to the provision of costly effort, standard economic theory suggests that monetary incentives could compensate for the effort undertaken, leading to a positive relationship between monetary incentives and the amount of effort expended. On the empirical side, decades of experimental research in psychology and economics has revealed that the relationship between monetary incentives and effort provision may not be as straightforward as predicted by standard economic theory, with some papers showing a positive relationship (Jenkins et al., 1998; DellaVigna and Pope, forthcoming), some showing no relationship (Araujo et al., 2016), and some showing a non-monotonic relationship (Gneezy and Rustichini, 2000; Pokorny, 2008; Ariely et al., 2009).

The study of the role of monetary incentives is complicated by the fact that there may be non-monetary incentives that influence behavior. This implies that in real-effort experiments, it may be difficult to detect the impact of monetary incentives if it is likely to be overshadowed by the presence of non-monetary incentives. Our aim in this paper, therefore, is to study the impact of monetary and non-monetary incentives by disentangling them. In the real world, monetary as well as non-monetary incentives contribute to effort in important ways. However, the laboratory gives us a chance to analyze the relationship between monetary incentives and effort by mitigating the impact of non-monetary incentives. We show a positive relationship exists between monetary incentives and effort provision, as predicted by standard economic theory.

In the literature on incentive effects, our paper is also related to the group of papers which study the interactions between different type of incentives (see e.g., Deci et al., 1999; Gneezy et al., 2011; Bowles and Polanía-Reyes, 2012; Kamenica, 2012). This literature distinguishes between extrinsic and intrinsic motivations and largely focuses on the crowding out of intrinsic motivations by extrinsic rewards. In contrast, focusing on real-effort tasks, we distinguish between monetary and non-monetary incentives. Monetary incentives appeal to an individual’s extrinsic motivation, while non-monetary incentives can appeal to an individual’s intrinsic (e.g., the inherent desire to win) or extrinsic (e.g., concerns about how one is perceived by others) motivation. Although both monetary and

\footnote{See Camerer and Hogarth (1999) for a survey. Takahashi et al. (2016) show that the nature of the task may influence how subjects respond to monetary incentives.}
non-monetary incentives co-exist in our Experiment 1, examining crowding-out effects is not the focus of our study.\footnote{The intrinsic vs. extrinsic motivation literature is largely concerned with crowding out of intrinsic motivations in domains such as prosocial activities, lifestyle habits, and education.}

Our second contribution is to the literature on contests. Unlike most of the studies which analyze monetary incentives using piece-rate schemes, we consider the role of monetary incentives in contests. Contests are used in many economic, political, and social environments. Hence, understanding optimal contest design (i.e., how effort depends on monetary versus non-monetary incentives in contests) has significant implications for social welfare. As a result, contest behavior has been studied extensively in the laboratory (see Dechenaux et al., 2015, for a survey). However, most of these studies involve a stated-effort design. An important question is whether the results obtained using stated-effort designs continue to hold with real-effort tasks. Our paper provides guidance on how real-effort tasks can be used in contests to study the relationship between incentives and effort. More specifically, we show in Experiment 1 that the positive relationship reported in studies with a stated-effort design does not necessarily hold with a real-effort task. As we show in Experiments 2 and 3, the specific roles that monetary and non-monetary incentives play in contests need to be considered more carefully.

Finally, our paper makes a methodological contribution by analyzing the impact of offering different type of outside options in real-effort tasks and comparing their effectiveness in mitigating non-monetary incentives. Several studies have considered outside options as additional features to the design of the tasks. For example, subjects have been provided with the option to leave the laboratory early (e.g., Dickinson, 1999; Rosaz et al., 2016), browse newspapers or magazines (e.g., Charness et al., 2014; Rey-Biel et al., 2017), surf the Internet (e.g., Kessler and Norton, 2016), watch pre-selected popular YouTube videos (e.g., Hayashi et al., 2013), press a paid timeout button (e.g., Mohnen et al., 2008; Blumkin et al., 2012), or participate in an alternative activity (e.g., van Dijk et al., 2001).\footnote{See Kurzban et al. (2013) for a related discussion in the psychology literature about the link between outside options and opportunity cost of effort.} However, a direct comparison of treatments with and without outside options in order to detect the impact of outside options is not the goal of these papers.
Two papers that are more closely related in this regard are Corgnet et al. (2015) and Eckartz (2014). Corgnet et al. (2015) study the impact of on-the-job leisure on work performance under individual versus team-production incentive schemes. They show that, on average, the availability of the Internet as an outside option has a significant and negative impact on effort under team pay but not under individual pay. Eckartz (2014) studies the impact of different compensation schemes (flat wage, tournament and piece rate) on effort provision. She finds that the availability of a paid pause button as an outside option decreases effort provision under the flat wage scheme, but it has no impact on effort provision under tournament or piece-rate incentives. Although the research objectives of these two papers are different from ours and they consider only one type of outside option, they provide complementary findings in that they also show the introduction of outside options can reduce effort provision.7

4.3 Experiment 1

4.3.1 Design and Treatments

Our first experiment aims to investigate whether individuals respond to changes in monetary incentives by altering their effort levels. Subjects participate in a contest to determine how two prizes (Prize 1 and Prize 2) will be allocated between themselves and their matched partner. Each subject i chooses an effort level, $e_i$. Subjects make their effort choice by performing a real-effort encryption task (Erkal et al., 2011) for 10 minutes. In this task, each subject is given an encryption table that assigns a number to each letter of the alphabet, and is asked to encrypt a pre-determined sequence of letter combinations by substituting them with numbers using the encryption table. Each combination in the sequence consists of five randomly generated letters, and all the subjects are given the same combinations of letters in the same sequence. Figure 4.1 shows the screenshot of the task for this treatment.

7See also Gächter et al. (2016), who introduce a novel real-effort task with an induced cost of effort and show that subjects respond to monetary incentives. Outside of the labor market, Lei et al. (2001) suggest that the occurrence of bubbles in asset markets in the laboratory may be partly due to the lack of alternative activities that subjects can participate in apart from trading in the asset market.
Figure 4.1: Screenshot for treatments B, BL, and BH

At the end of 10 minutes, the number of letter combinations that a subject successfully encrypts is defined as his/her effort choice, $e_i$. The probability of winning Prize 1, $p_i(e_i, e_j)$, is given by

$$p_i(e_i, e_j) = \frac{e_i}{e_i + e_j},$$

where $e_j$ is the effort level chosen by their matched partner. The subject who does not receive Prize 1 receives Prize 2. If both subjects choose an effort level of zero, then the prizes are allocated randomly.

We vary the monetary incentives by adjusting the value of Prize 1 across three different treatments. Panel (a) of Table 4.1 provides a summary of the treatments for this experiment. In the Baseline (B) treatment, both Prize 1 and Prize 2 are valued at 150 Experimental Currency Units (ECU). In this treatment, there are no monetary incentives for doing the task since both prizes are equal to each other, and each subject is guaranteed to receive one of the two prizes. In the Baseline – Low (BL) treatment, the value of Prize 1 is increased to 250 ECU, while the value of Prize 2 remains at 150 ECU. In the Baseline – High (BH) treatment, the value of Prize 1 is further increased to 600 ECU. Hence, treatments BL and BH represent situations where there are low and high
monetary incentives for doing the task, respectively.

We consider two key variables of interest. First, we are interested in the participation rate, which is the proportion of subjects who exert positive effort. In our analysis, Participation is a dummy variable that takes the value of 1 if a subject exerts positive effort. Second, we are interested in the level of effort per minute, which is defined for those subjects who choose to exert a positive effort level as the average number of letter combinations encrypted per minute. Taken together, these two variables are analogous to the extensive and intensive margins of labor supply commonly used in the labor literature.

4.3.2 Hypotheses

In the absence of any monetary incentives in treatment B, non-monetary incentives (discussed in more detail in Section 4.1) are the key determinants of effort. If the non-monetary incentives are sufficiently strong and higher than the cost of effort, then subjects may choose to exert effort even though both Prize 1 and Prize 2 are equal. In fact, evidence from the experimental literature suggests that subjects tend to exert high amounts of effort in real-effort tasks even though there are no additional monetary incentives to be gained from doing so (see, e.g., Kuhnen and Tymula, 2012; Charness et al., 2014; Masclet et al., 2015). In line with these previous studies, we conjecture that a significant proportion of subjects will exert positive effort in treatment B.

**Hypothesis 4.1:** In the absence of monetary incentives in treatment B, subjects will be motivated by non-monetary incentives in their effort choice. Hence, in treatment B, both the participation rate and average effort per minute will be greater than zero.

When the value of Prize 1 is increased relative to Prize 2 in treatments BL and BH, monetary incentives will also be driving behavior. This is because subjects can now receive a monetary gain by exerting effort to increase their probability of receiving the higher prize. Consequently, we expect effort provision to be the lowest in treatment B and to increase as the value of Prize 1 increases.

---

8Similar results are reported in the psychology literature (see, e.g. Huxtable et al., 1946; Orne, 1962) which show that subjects exert effort for many hours (even days) with insignificant amounts of monetary incentives.
### Table 4.1: Summary of experimental treatments

<table>
<thead>
<tr>
<th>Treatment</th>
<th># subjects</th>
<th>Prize 1 (ECU)</th>
<th>Prize 2 (ECU)</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Experiment 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline (B)</td>
<td>54</td>
<td>150</td>
<td>150</td>
<td>Baseline treatment with no incentives</td>
</tr>
<tr>
<td>Baseline – Low (BL)</td>
<td>50</td>
<td>250</td>
<td>150</td>
<td>Baseline treatment with low incentives</td>
</tr>
<tr>
<td>Baseline – High (BH)</td>
<td>52</td>
<td>600</td>
<td>150</td>
<td>Baseline treatment with high incentives</td>
</tr>
<tr>
<td>(b) Experiment 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline (B)</td>
<td>54</td>
<td>150</td>
<td>150</td>
<td>Baseline treatment</td>
</tr>
<tr>
<td>Early Departure (ED)</td>
<td>50</td>
<td>150</td>
<td>150</td>
<td>Outside option: Can leave experiment early</td>
</tr>
<tr>
<td>Pause (P)</td>
<td>56</td>
<td>150</td>
<td>150</td>
<td>Outside option: Timeout payment</td>
</tr>
<tr>
<td>Activity 1/2 (A1A2)</td>
<td>56</td>
<td>150</td>
<td>150</td>
<td>Outside option: Same activity with piece rate</td>
</tr>
<tr>
<td>(c) Experiment 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Activity 1/2 (A1A2)</td>
<td>56</td>
<td>150</td>
<td>150</td>
<td>A1A2 treatment with no incentives</td>
</tr>
<tr>
<td>Activity 1/2 – Low (A1A2-L)</td>
<td>56</td>
<td>250</td>
<td>150</td>
<td>A1A2 treatment with low incentives</td>
</tr>
<tr>
<td>Activity 1/2 – High (A1A2-H)</td>
<td>52</td>
<td>600</td>
<td>150</td>
<td>A1A2 treatment with high incentives</td>
</tr>
</tbody>
</table>
**Hypothesis 4.2**: The presence of monetary incentives in addition to non-monetary incentives in treatments BL and BH will lead to higher participation and effort per minute in these treatments as compared to treatment B. Moreover, increasing monetary incentives will result in an increase in both the participation rate and effort per minute in treatment BH as compared to treatment BL.

### 4.3.3 Procedures

Our experiments were conducted between August and November 2015 in the Experimental Economics Laboratory at the University of Melbourne (E2MU) and programmed using z-Tree (Fischbacher, 2007). Subjects were recruited using ORSEE (Greiner, 2015) and consisted mostly of students from the University of Melbourne. Our experiment follows a between-subject treatment design where subjects were randomly assigned into one of the three treatments. We ran two sessions for each treatment, with 24-28 subjects in each session. Each session lasted between 60 and 90 minutes, including time for reading the instructions and the administration of payments. We collected 156 independent observations in total for Experiment 1.

Upon entering the laboratory, subjects first participated individually in a one-shot risk task (Gneezy and Potters, 1997). Each subject was given an endowment of 50 ECU and had to decide how much of this endowment to invest in a risky asset. With equal probabilities, the asset either returned three times the amount invested or nothing. Subjects got to keep any part of their endowment that was not invested.

After completing the risk task, subjects received the instructions for the contest task.\(^9\) Subjects were randomly divided into pairs, and they did not know the identity of their matched partners. Instructions were summarized verbally by the experimenter, and subjects participated in a practice real-effort task for three minutes to allow them to familiarize themselves with the interface of the task.

At the end of the experiment, subjects were asked to answer a questionnaire which included basic demographic questions as well as questions about their decisions during

\(^9\)Instructions are available in Appendix C.1.
the experiment. Once they completed the questionnaire, subjects were paid their earnings privately. Subjects were randomly paid for their decisions in either the risk task or the contest task, and this was determined at the session level. Subjects did not learn the outcomes of the tasks until the end of the experiment. Earnings were converted to Australian dollars (AUD) at the rate 10 ECU = 1 AUD and included a $10 show-up fee. Subjects earned between $10 and $70, with the mean earnings being $22.59.

4.3.4 Results

Figure 4.2 presents histograms of the distribution of effort in treatments B, BL, and BH. Overall, the histograms show that: (i) a large number of subjects choose to exert positive effort in all three treatments; and (ii) the distribution of effort is almost identical across the three treatments. A Kruskal-Wallis equality-of-populations rank test fails to reject the null hypothesis that the distribution of effort in treatments B, BL, and BH are equal to one another (p-value = 0.241). Pairwise Kolmogorov-Smirnov tests for the equality of distribution functions also fail to reject the null hypotheses that the distributions of effort between any pair of treatments are equal to each other (B vs. BL: p-value = 0.296; B vs. BH: p-value = 0.685; BL vs. BH: p-value = 0.831).

Panel (a) of Table 4.2 presents the participation rate and average effort per minute conditional on participating. Overall, almost all of the subjects (96%) exert positive effort in the absence of monetary incentives in treatment B. Conditional on participation, these subjects also exert a non-trivial amount of effort, with the average effort per minute being 4.22. Hence, we find support for Hypothesis 4.1 in that subjects are motivated by non-monetary incentives in their effort choice and choose to exert high effort even in the absence of monetary incentives.

One may argue that this result is sensitive to the length and type of the specific task used in our experiment. Fatigue and boredom may play more important roles if the task is longer or different. Hence, in such environments, the impact of non-monetary incentives

---

This payment mechanism is commonly used in the experimental literature. Bardsley et al. (2010) show that it is incentive compatible under the assumption of expected utility maximization. They also discuss the validity of using this mechanism when agents are not expected utility maximizers.
Table 4.2: Summary statistics

<table>
<thead>
<tr>
<th>Treatment</th>
<th># subjects&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Participation rate</th>
<th>Average effort per minute&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Experiment 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline (B)</td>
<td>54</td>
<td>0.96</td>
<td>4.22</td>
</tr>
<tr>
<td>Baseline – Low (BL)</td>
<td>50</td>
<td>1.00</td>
<td>4.49</td>
</tr>
<tr>
<td>Baseline – High (BH)</td>
<td>52</td>
<td>1.00</td>
<td>4.33</td>
</tr>
<tr>
<td>(b) Experiment 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline (B)</td>
<td>54</td>
<td>0.96</td>
<td>4.22</td>
</tr>
<tr>
<td>Early Departure (ED)</td>
<td>50</td>
<td>0.82</td>
<td>3.27</td>
</tr>
<tr>
<td>Pause (P)</td>
<td>56</td>
<td>0.54</td>
<td>0.94</td>
</tr>
<tr>
<td>Activity 1/2 (A1A2)</td>
<td>53</td>
<td>0.38</td>
<td>0.54</td>
</tr>
<tr>
<td>(c) Experiment 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Activity 1/2 (A1A2)</td>
<td>53</td>
<td>0.38</td>
<td>0.54</td>
</tr>
<tr>
<td>Activity 1/2 – Low (A1A2-L)</td>
<td>56</td>
<td>0.98</td>
<td>2.52</td>
</tr>
<tr>
<td>Activity 1/2 – High (A1A2-H)</td>
<td>52</td>
<td>0.96</td>
<td>3.27</td>
</tr>
</tbody>
</table>

Sample means given. Standard deviations in brackets.
For treatments A1A2, A1A2-L, and A1A2-H, we report the participation rate and average effort per minute in Activity 1.

<sup>a</sup> We discovered that three subjects in treatment A1A2 had previously participated in the same experiment. Hence, we drop these subjects in our analysis.

<sup>b</sup> Average effort per minute is defined as the average effort by subjects who exerted a positive level of effort.
may be weaker. To test this, we ran two variants of treatment B. In the 20-Minute (20-Min) treatment, subjects performed the encryption task for 20 minutes instead of 10 minutes ($N = 56$). In the Slider (S) treatment, the subjects participated in a slider task (Gill and Prowse, 2012) instead of the encryption task ($N = 54$). We find no statistically significant differences in behavior between treatment B and these two treatments.$^{11}$

Panel (a) of Table 4.2 also reveals that when monetary incentives are present, 100% of subjects exert positive effort in both of the treatments (BL and BH). Pairwise Fisher’s exact tests fail to reject the null hypotheses that the participation rates are equal between the treatments with monetary incentives and the baseline treatment (B vs. BL: p-value

$^{11}$The details of these treatments can be found in Appendix C.2. It is possible that considering even longer tasks may create a bigger opportunity for monetary incentives to affect behavior. Further exploration of this issue is beyond the scope of the current paper. Our design choice is guided by the literature where most experiments do not include such long tasks.
= 0.496; B vs. BH: p-value = 0.495). Conditional on participating, the average effort per minute is 4.49 in treatment BL and 4.33 in treatment BH. Pairwise Wilcoxon rank-sum tests fail to reject the null hypotheses that the effort per minute between any pair of treatments are equal to each other (B vs. BL: p-value = 0.172; B vs. BH: p-value = 0.421; BL vs. BH: p-value = 0.601).

Table 4.3 presents the coefficient estimates from OLS regressions of participation (column 1) and effort per minute (column 2) on treatment variables and subjects’ characteristics, which include the percentage of the endowment invested in the risk task, age, gender, study level at University, whether the subject is pursuing a major in economics, whether the subject is Australian, and previous experience with economic experiments. The table also presents the test of equality between the estimated coefficients for treatments BL and BH. Treatment B is the reference treatment for both regressions.

The regression estimates in Table 4.3 are consistent with the conclusions from the non-parametric tests, with the exception of the estimated coefficient on treatment BL in column (2). The estimated coefficient reveals that, conditional on participating, effort per minute is 0.329 higher on average when low monetary incentives are present as compared to when no monetary incentives are present. This effect is statistically significant at the 5% level (p-value = 0.015). To investigate this issue further, column (3) presents the results for median effort using a quantile regression. Consistent with the conclusions from our non-parametric tests, the coefficient estimates indicate that introducing monetary incentives (treatments B vs. BL) and increasing monetary incentives (treatments BL vs. BH) have no significant impact on the median effort.12

In summary, consistent with the results from the literature, we fail to find systematic support for Hypothesis 4.2. The regression estimates using average effort suggest that introducing monetary incentives may have an impact on effort per minute. However,

12Given that participation is similar across the three treatments in Experiment 1, total performance is another measure that can be used to analyze the data. The results for unconditional effort per minute (total effort divided by the total time spent on the task) are qualitatively similar to those for conditional effort per minute, with the exception that the difference between treatments B and BL is statistically significant. This result seems to be driven by the two subjects in treatment B who exert zero effort and disappears when we consider quantile (median) regressions. The results for unconditional and conditional effort are similar in Experiments 2 and 3.
Table 4.3: Regression results for Experiment 1

<table>
<thead>
<tr>
<th>Variables</th>
<th>Participation =1 if Effort &gt;0 given Effort &gt;0</th>
<th>Effort per minute given Effort &gt;0</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(OLS)</td>
<td>(Quantile)</td>
</tr>
<tr>
<td>Baseline – Low (BL)</td>
<td>0.032</td>
<td>0.329**</td>
</tr>
<tr>
<td></td>
<td>(0.023)</td>
<td>(0.133)</td>
</tr>
<tr>
<td>Baseline – High (BH)</td>
<td>0.035</td>
<td>0.145</td>
</tr>
<tr>
<td></td>
<td>(0.022)</td>
<td>(0.131)</td>
</tr>
<tr>
<td>% invested in risk task</td>
<td>-0.059</td>
<td>-0.296</td>
</tr>
<tr>
<td></td>
<td>(0.039)</td>
<td>(0.233)</td>
</tr>
<tr>
<td>Age</td>
<td>-0.004</td>
<td>-0.030*</td>
</tr>
<tr>
<td></td>
<td>(0.003)</td>
<td>(0.016)</td>
</tr>
<tr>
<td>Female</td>
<td>-0.003</td>
<td>0.068</td>
</tr>
<tr>
<td></td>
<td>(0.019)</td>
<td>(0.114)</td>
</tr>
<tr>
<td>Economics major</td>
<td>0.006</td>
<td>-0.938***</td>
</tr>
<tr>
<td></td>
<td>(0.036)</td>
<td>(0.212)</td>
</tr>
<tr>
<td>Undergraduate student</td>
<td>-0.063</td>
<td>1.854***</td>
</tr>
<tr>
<td></td>
<td>(0.098)</td>
<td>(0.578)</td>
</tr>
<tr>
<td>Graduate student</td>
<td>-0.045</td>
<td>2.094***</td>
</tr>
<tr>
<td></td>
<td>(0.090)</td>
<td>(0.528)</td>
</tr>
<tr>
<td>Australian</td>
<td>-0.031</td>
<td>0.188</td>
</tr>
<tr>
<td></td>
<td>(0.020)</td>
<td>(0.116)</td>
</tr>
<tr>
<td># previous experiments</td>
<td>0.003</td>
<td>0.028</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.031)</td>
</tr>
<tr>
<td>Constant</td>
<td>1.166***</td>
<td>3.060***</td>
</tr>
<tr>
<td></td>
<td>(0.137)</td>
<td>(0.814)</td>
</tr>
<tr>
<td>Test:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BH = BL</td>
<td>0.003</td>
<td>-0.184</td>
</tr>
<tr>
<td></td>
<td>(0.023)</td>
<td>(0.134)</td>
</tr>
<tr>
<td>Observations</td>
<td>156</td>
<td>154</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.083</td>
<td>0.308</td>
</tr>
</tbody>
</table>

Ordinary least squares estimates given in (1) and (2), and quantile (median) regression estimates given in (3). Standard errors in parentheses. For all regressions, treatment B is the reference treatment. In all regressions, the controls are treatment variables and subjects’ characteristics, which include the percentage of the endowment invested in the risk task, age, gender, study level at University, whether the subject is pursuing a major in economics, whether the subject is Australian, and previous experience with economic experiments.

*** p<0.01, ** p<0.05, * p<0.10.
this effect is not observed with the non-parametric analysis or when we consider median effort. Increasing monetary incentives from treatment BL to treatment BH does not have a statistically significant impact on the participation rate and effort per minute either.

The results from Experiment 1 suggest that non-monetary incentives are present and lead to high effort even in the absence of monetary incentives. This may make it difficult to capture whether subjects respond to monetary incentives in real-effort tasks within the laboratory. This is problematic if our goal is to study the responsiveness of effort to monetary incentives. Hence, we next explore whether we can detect the impact for monetary incentives if we reduce the impact of non-monetary incentives.

4.4 Experiment 2

4.4.1 Framework

Our design in Experiment 2 is motivated by the following simple framework, where subjects participating in real-effort tasks are assumed to make their effort choices by comparing the marginal cost and marginal benefit of exerting one more unit of effort. Formally, let $e_i$ denote the effort chosen by subject $i$. We assume that the marginal cost of effort is determined by the opportunity cost of time and is convex in effort ($ce_i^2$). For example, the opportunity cost of time in Experiment 1 is determined by sitting idle and daydreaming.\textsuperscript{13}

On the benefit side, let $V_1$ and $V_2$ represent the value of Prize 1 and Prize 2, respectively. Hence, $V_2 + \frac{e_i}{e_i + e_j}(V_1 - V_2)$ is the monetary payoff subjects expect to receive in a two-player contest. In addition to monetary incentives, subjects’ behavior may also be influenced by: (i) curiosity for the task or enjoyment of the task; (ii) the utility derived from doing what is expected of them or being useful/productive (experimenter demand effect);\textsuperscript{14} and (iii) the utility derived from being the winner in a competitive environment.\textsuperscript{15} Assuming concavity, we let $b_1 \log(e_i)$ and $b_2 \log(e_i)$ stand for (i) and (ii),

\textsuperscript{13} Subjects were not allowed to have access to their phones or books during the experiment.

\textsuperscript{14} Experimenter demand effect refers to the expectations that subjects form about their effort/performance based on the cues that they receive from the environment created by the experimenter (e.g., from the experimenter’s instructions). Note that it is also possible for their expectations to be shaped by other factors, such as the goals they may set for themselves.

\textsuperscript{15} These non-monetary incentives have been found to influence subjects’ behavior in other environments.
respectively. Since we do not have public announcement of winners in our design, (iii) corresponds to the utility that subjects may derive from privately receiving the title of “recipient of Prize 1” at the end of the experiment. Given that a subject receives this benefit only if s/he wins the contest, \( b_3 \frac{e_i}{e_i + e_j} \) stands for (iii).

Putting these together, individual \( i \) solves

\[
\max_{e_i} V_2 + \frac{e_i}{e_i + e_j} (V_1 - V_2) + b_1 \log(e_i) + b_2 \log(e_i) + b_3 \frac{e_i}{e_i + e_j} - c e_i^2. \tag{4.1}
\]

Since we are interested in studying non-monetary incentives, we remove the monetary incentives in Experiment 2 by setting \( V_1 = V_2 = 150 \) ECU. Then, the following first-order condition (FOC) defines the optimal effort of subject \( i \):

\[
\frac{b_1 + b_2}{e_i} + \frac{e_j}{(e_i + e_j)^2} b_3 = 2c e_i. \tag{4.2}
\]

From equation (4.2), it is clear that effort can be reduced either by decreasing \( b_1, b_2, \) and \( b_3 \), or by increasing \( c \).

### 4.4.2 Treatments and Hypotheses

Panel (b) of Table 4.1 presents a summary of the three treatments designed for this purpose. All three treatments follow the same setup as in Experiment 1 with the modifications we explain below. Treatment B from Experiment 1 serves as the baseline treatment for comparison in this experiment.

In Experiment 1, the opportunity cost of effort, \( c \), depends on the utility derived from sitting idle. The first two treatments in Experiment 2 aim to increase \( c \) in equation (4.2). Specifically, in the Early Departure (ED) treatment, subjects are permitted to leave the experiment at any time during the task. If they wish to leave the experiment early, they also (see, e.g., Carpenter et al., 2010a; Sheremeta, 2010; Zizzo, 2010). In a related paper, DellaVigna and Pope (forthcoming) consider other types of non-monetary motivators of behavior, namely social preferences, present bias, reference dependence, social comparison, and task significance.

\(^{16}\)The formulation we present assumes that monetary and non-monetary incentives are additively separable as in Frey and Oberholzer-Gee (1997) and DellaVigna and Pope (forthcoming). See Bowles and Polania-Reyes (2012) for an alternative, non-separable formulation.
can click a button on the screen to stop the task. In this case, they still receive 150 ECU for this task since both prizes are of the same monetary value. Once they stop the task, these subjects are invited to complete the questionnaire, and are then paid their earnings in private before leaving the laboratory.

In the Pause (P) treatment, subjects are permitted to take timeout at any time during the real-effort task and as many times as they wish. To take timeout, they have to click a button on the screen to pause the task temporarily. Subjects receive 5 ECU per minute they spend in timeout which is paid on a pro rata basis. While taking timeout, subjects can click another button on the screen to return to the task at any time.

In summary, in treatment ED, the opportunity cost of effort is given by the value of the time spent doing something outside of the laboratory (e.g., enjoying the company of one’s friends). In treatment P, the opportunity cost of effort has an explicit monetary value, 5 ECU per minute. The values of these outside options are likely to be higher than the value of sitting idle. Our hypothesis regarding these two treatments is the following.

**Hypothesis 4.3:** Allowing subjects to leave the experiment early and/or introducing an incentivized timeout button will increase the opportunity cost of effort. Hence, relative to treatment B, both the participation rate and average effort per minute will be lower in treatments ED and P.

The third treatment in Experiment 2 is designed to increase $c$ (the opportunity cost of effort), reduce $b_1$ (task enjoyment), and reduce $b_2$ (experimenter demand effect) at the same time. To this end, in the Activity 1/2 (A1A2) treatment, subjects are provided with two activities simultaneously on the same screen. Figure 4.3 shows the screenshot of the task for this treatment. Activity 1 is located on the top half of the screen while Activity 2 is located on the bottom half. In both activities, subjects participate in the

\[17\] Eckartz (2014) and Mohnen et al. (2008) also use an incentivized timeout button as an outside option in their design, while Dickinson (1999) uses early departure to study on- and off-the-job leisure. We considered treatments with other outside options as well, such as newspapers/magazines or the Internet. However, results from our pilot experiments suggest that, within the timeframe provided for the real-effort task (10 minutes), these outside options do not significantly influence subjects’ behavior in the laboratory. Hence, we decided not to proceed with these treatments.
encryption task, but with different sets of letter combinations and encryption tables. The payment scheme for Activity 1 is exactly the same as in treatment B, where the number of letter combinations encrypted by a pair of subjects determines how Prize 1 and Prize 2, which are equal to each other, will be allocated between them. In Activity 2, subjects participate in the task individually and are paid a piece rate of 1.25 ECU for each letter combination encrypted.\textsuperscript{18} Subjects can choose to do only one, both, or none of the activities, and they can switch between the two activities at any time during the task and as many times as they want. Their payment from this part of the experiment is the sum of the payments received from Activity 1 and Activity 2.\textsuperscript{19}

In treatment A1A2, conditional on choosing to exert effort on the task, subjects decide how to allocate their effort between the two activities. We are mainly interested in the effort exerted in Activity 1 since Activity 2 is considered to be the outside option. More specifically, the decision to put in effort in Activity 2 is similar to taking a timeout (as in treatment P), except that the subject now has to work to receive the timeout payment. As a result, the piece rate earned in Activity 2 is the opportunity cost of exerting effort in Activity 1. Moreover, since Activity 2 is based on the same task as Activity 1, we expect that task enjoyment and experimenter demand effect will no longer be the drivers of effort in Activity 1. This is because subjects can derive the same utility from these two factors in both Activity 1 and Activity 2. Hence, we would expect only those subjects who are sufficiently competitive to exert effort in Activity 1.

Formally, subject $i$ in treatment A1A2 solves

$$\max_{e_{1,i}, e_{2,i}} V_2 + pe_{2,i} + (b_1 + b_2) \log(e_{1,i} + e_{2,i}) + b_3 \frac{e_{1,i}}{e_{1,i} + e_{1,j}} - c(e_{1,i} + e_{2,i})^2$$

(4.3)

where $e_{1,i}$ and $e_{2,i}$ are the effort exerted by individual $i$ in Activity 1 and Activity 2.

\textsuperscript{18}We calibrate the piece rate such that, if a subject spends all his/her time on Activity 2, then s/he will receive 50 ECU from Activity 2 on average, which is the same as the total timeout payment available in treatment P. This calibration is done using the average number of letter combinations encrypted by subjects within 10 minutes in our pilot experiments.

\textsuperscript{19}Similarly, van Dijk et al. (2001) consider a design with two activities to investigate the impact of different incentive schemes. Johnson and Salmon (2016) also introduce an alternative activity (a counting task) as an outside option to their main real-effort task (a number adding task). Since the two activities are different, subjects may continue to be motivated by task enjoyment or curiosity in their effort decisions.
respectively, and \( \rho \) is the piece rate paid per unit of effort exerted in Activity 2. The FOCs with respect to \( e_{1,i} \) and \( e_{2,i} \) are given by

\[
\frac{b_1 + b_2}{e_{1,i} + e_{2,i}} + b_3 \frac{e_{1,j}}{(e_{1,i} + e_{1,j})^2} = 2c(e_{1,i} + e_{2,i})
\]

and

\[
\frac{b_1 + b_2}{e_{1,i} + e_{2,i}} + \rho = 2c(e_{1,i} + e_{2,i}),
\]

respectively. These two equations, when combined, give us

\[
\frac{b_3}{(e_{1,i} + e_{1,j})^2} = \rho. \tag{4.4}
\]

Equation (4.4) implies that subject \( i \)'s effort allocation will be such that on the margin,
the benefit derived from being the recipient of Prize 1 in Activity 1 is equal to the piece rate received in Activity 2. Hence, \(b_1\) and \(b_2\) no longer play a role in the allocation of effort between Activity 1 and Activity 2. However, if \(b_3\) is sufficiently high, then we would expect subjects to exert a positive level of effort in Activity 1. Nevertheless, the average participation and effort levels in Activity 1 should be lower than those in treatment B. This leads us to the following hypothesis for treatment A1A2.

**Hypothesis 4.4:**

(i) Providing an alternative activity that pays a piece rate will increase the opportunity cost of effort and remove the impact of task enjoyment and experimenter demand effect on effort provision in Activity 1 of treatment A1A2. Hence, relative to treatment B, both the participation rate and average effort per minute will be lower in Activity 1 of treatment A1A2.

(ii) Since subjects may still be motivated by a desire to be the recipient of Prize 1 in the contest (competitiveness), both the participation rate and average effort per minute in Activity 1 of treatment A1A2 will be greater than zero.

The treatments for Experiment 2 were also conducted in the \(E^2MU\) lab, and they followed the same procedures as in Experiment 1. We ran two sessions each for treatments ED, P, and A1A2, and collected a total of 162 independent observations in these treatments. Subjects earned between $10 and $32.90, with the mean earnings being $22.56.

### 4.4.3 Results

Figure 4.4 shows the histograms of the subjects’ effort choices by treatment. While the majority of subjects exert high effort in the baseline treatment B, this is no longer the case in treatments ED, P, and A1A2. Pairwise Kolmogorov-Smirnov tests reject the null hypotheses that the distribution of effort in treatments ED, P, and A1A2, are equal to that in treatment B (B vs. ED: p-value < 0.001; B vs. P: p-value < 0.001; B vs. A1A2: p-value < 0.001).
Panel (b) of Table 4.2 presents summary statistics at the treatment level. The participation rates in treatments ED (82%), P (54%), and A1A2 (38%) are lower than that in the baseline treatment B (96%). Pairwise Fisher’s exact tests reject the null hypotheses that the participation rates in these treatments are equal to that in treatment B (B vs. ED: p-value = 0.025; B vs. P: p-value < 0.001; B vs. A1A2: p-value < 0.001). Conditional on participation, the average effort per minute is 3.27 in treatment ED, 0.94 in treatment P, and 0.54 in treatment A1A2, all of which are lower than that in treatment B (4.22). Pairwise Wilcoxon rank-sum tests reject the null hypotheses that the effort per minute in these treatments are equal to that in treatment B (B vs. ED: p-value < 0.001; B vs. P: p-value < 0.001; B vs. A1A2: p-value < 0.001). Overall, the results suggest that these treatments are effective in reducing effort on both the extensive and intensive
margins.

Table 4.4 presents the results of OLS and quantile (median) regressions of participation (column 1) and effort per minute (columns 2 and 3). For all regressions, Treatment B is the reference treatment. The regression estimates for treatments ED, P and A1A2 are consistent with the conclusions from the non-parametric tests. Both the participation rate and average effort per minute in these treatments are significantly lower than those in treatment B. Hence, we find support for Hypotheses 4.3 and 4.4(i).

Panel (b) of Table 4.2 also reveals that a non-trivial proportion of the subjects (38%) exert positive effort in Activity 1 of treatment A1A2. Conditional on participation, the average effort per minute is 0.54, suggesting that subjects exert a small amount of effort so that they have a positive probability of receiving Prize 1. Hence, we conclude that, in support of Hypothesis 4.4(ii), despite the high opportunity cost of exerting effort in Activity 1 (given by the piece rate in Activity 2) and the lack of monetary incentives, some subjects still choose to exert effort in this activity.\textsuperscript{20}

We next explore which of the treatments ED, P, and A1A2 is more effective in mitigating the impact of non-monetary incentives. Pairwise Kolmogorov-Smirnov tests reject the null hypotheses that the distribution of effort in treatments ED, P, and A1A2 are equal to one another (ED vs. P: p-value < 0.001; ED vs. A1A2: p-value < 0.001; P vs. A1A2: p-value = 0.060). From panel (b) of Table 4.2, we observe that among these three treatments, both the participation rate and average effort per minute are the highest in treatment ED and the lowest in treatment A1A2. Pairwise Fisher’s exact tests and Wilcoxon rank-sum tests reveal that both the participation rate and effort per minute in treatment ED are significantly higher than those in treatments P and A1A2 (Pairwise Fisher’s exact tests: ED vs. P: p-value = 0.002, ED vs. A1A2: p-value < 0.001; Pairwise Wilcoxon rank-sum tests: ED vs. P: p-value < 0.001, ED vs. A1A2: p-value < 0.001). Hence, of these three treatments, treatment ED is the least effective method for

\textsuperscript{20}Competitiveness can affect behavior in a contest even in the absence of a real-effort task. In an additional treatment that we ran with stated effort (treatment EC, N = 56), where we had the same parameters for Prize 1 and Prize 2 as in treatment B, we find that the behavior of the subjects is largely similar to that in Activity 1 of treatment A1A2. This finding is consistent with Sheremeta (2010). The details can be found in Online Appendix C.2.
### Table 4.4: Regression results for Experiment 2

<table>
<thead>
<tr>
<th>Variables</th>
<th>Participation =1 if Effort &gt;0 (OLS)</th>
<th>Effort per minute given Effort &gt;0 (OLS)</th>
<th>Effort per minute given Effort &gt;0 (Quantile)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early Departure (ED)</td>
<td>-0.184** (0.084)</td>
<td>-0.979*** (0.255)</td>
<td>-0.668*** (0.196)</td>
</tr>
<tr>
<td>Pause (P)</td>
<td>-0.438*** (0.078)</td>
<td>-3.260*** (0.268)</td>
<td>-3.863*** (0.206)</td>
</tr>
<tr>
<td>Activity 1/2 (A1A2)</td>
<td>-0.596*** (0.080)</td>
<td>-3.778*** (0.318)</td>
<td>-4.111*** (0.244)</td>
</tr>
<tr>
<td>% invested in risk task</td>
<td>-0.226** (0.107)</td>
<td>-0.559 (0.382)</td>
<td>-0.063 (0.293)</td>
</tr>
<tr>
<td>Age</td>
<td>0.001 (0.007)</td>
<td>0.016 (0.022)</td>
<td>0.006 (0.017)</td>
</tr>
<tr>
<td>Female</td>
<td>-0.045 (0.059)</td>
<td>0.268 (0.209)</td>
<td>0.083 (0.161)</td>
</tr>
<tr>
<td>Economics major</td>
<td>-0.003 (0.092)</td>
<td>0.040 (0.331)</td>
<td>-0.169 (0.254)</td>
</tr>
<tr>
<td>Undergraduate student</td>
<td>0.038 (0.189)</td>
<td>1.008 (0.691)</td>
<td>0.759 (0.530)</td>
</tr>
<tr>
<td>Graduate student</td>
<td>0.104 (0.182)</td>
<td>1.050 (0.664)</td>
<td>0.651 (0.509)</td>
</tr>
<tr>
<td>Australian</td>
<td>-0.150** (0.061)</td>
<td>0.111 (0.225)</td>
<td>0.000 (0.173)</td>
</tr>
<tr>
<td># previous experiments</td>
<td>0.008 (0.016)</td>
<td>0.048 (0.052)</td>
<td>0.000 (0.040)</td>
</tr>
<tr>
<td>Constant</td>
<td>1.091*** (0.281)</td>
<td>2.951*** (0.957)</td>
<td>3.448*** (0.733)</td>
</tr>
<tr>
<td>Observations</td>
<td>213</td>
<td>143</td>
<td>143</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.290</td>
<td>0.651</td>
<td>-</td>
</tr>
</tbody>
</table>

Ordinary least squares estimates given in (1) and (2), and quantile (median) regression estimates given in (3). Standard errors in parentheses. For all regressions, treatment B is the reference treatment.

In all regressions, the controls are treatment variables and subjects’ characteristics, which include the percentage of the endowment invested in the risk task, age, gender, study level at University, whether the subject is pursuing a major in economics, whether the subject is Australian, and previous experience with economic experiments.

For treatment A1A2, we consider participation and effort per minute in Activity 1.

*** p<0.01, ** p<0.05, * p<0.10.
mitigating the impact of non-monetary incentives.

A comparison between treatments A1A2 and P reveals that the participation rates between these two treatments are not significantly different from each other (Fisher’s exact test: P vs. A1A2: p-value = 0.125). However, conditional on participating, the effort per minute in Activity 1 of treatment A1A2 is statistically significantly lower than that in treatment P (Wilcoxon rank-sum test: p-value = 0.026). Hence, while providing subjects with an alternative activity that pays a piece rate is no more effective in reducing participation than introducing an incentivized timeout button, it is the most effective method in reducing subjects’ effort level conditional on participating.

4.5 Experiment 3

4.5.1 Treatments and Hypotheses

Experiment 3, similar to Experiment 1, aims to detect responses to changes in monetary incentives. However, we now mitigate the impact of non-monetary incentives by using the insights from Experiment 2. Since we find treatment A1A2 to be the most effective, we choose to implement this treatment in Experiment 3 and vary monetary incentives.

Panel (c) of Table 4.1 provides a summary of the treatments in Experiment 3. In treatments A1A2-L and A1A2-H, we vary the value of Prize 1. In parallel with treatments BL and BH, Prize 1 is 250 ECU in treatment A1A2-L and 600 ECU in treatment A1A2-H. Hence, equation (4.4) becomes

\[
\frac{e_{1,j}}{(e_{1,i} + e_{1,j})^2} (V_1 - V_2) + b_3 \frac{e_{1,j}}{(e_{1,i} + e_{1,j})^2} = \rho. \tag{4.5}
\]

Since the impact of non-monetary incentives is reduced in these treatments, we expect monetary incentives to have a significant impact on participation and effort per minute in Activity 1.\textsuperscript{21}

\textsuperscript{21}Note that, despite the introduction of monetary incentives (which means subjects can now earn more by investing in Activity 1), subjects may still choose Activity 2 (which pays a piece rate) over Activity 1 if they have low confidence in their ability or if they are concerned about creating unequal payments (social preferences). Hence, any impact we find can be considered as a lower bound on incentive effects.
Hypothesis 4.5: Both participation and effort per minute will increase monotonically as the value of Prize 1 increases from 150 ECU (treatment A1A2) to 250 ECU (treatment A1A2-L) and 600 ECU (treatment A1A2-H).

Treatments A1A2-L and A1A2-H were also conducted in the E2MU lab, and they followed the same procedures as in Experiments 1 and 2. As before, we ran two sessions for each treatment and collected a total of 108 independent observations in these two treatments. Subjects earned between $10 and $75.40, with the mean earnings being $28.81.

4.5.2 Results

Figure 4.5 presents histograms of the distribution of effort in treatments A1A2, A1A2-L, and A1A2-H. A comparison of Figures 4.2 and 4.5 reveals that the distribution of effort in these three treatments are in stark contrast to the distribution of effort in treatments B, BL, and BH in Experiment 1. In the absence of monetary incentives (A1A2), the majority of the subjects exert zero or close to zero effort. When low incentives are present (A1A2-L), almost all the subjects appear to exert positive effort, although the distribution is still slightly skewed to the right. When high incentives are present (A1A2-H), the distribution of effort is now skewed to the left such that more than half of the subjects are exerting effort levels of 30 and above. A Kruskal-Wallis equality-of-populations rank test rejects the null hypothesis that the distribution of effort in treatments A1A2, A1A2-L, and A1A2-H are equal to one another (p-value = < 0.001). Pairwise Kolmogorov-Smirnov tests also reject the null hypotheses that the distributions of effort between any pair of treatments are equal to each other (A1A2 vs. A1A2-L: p-value = < 0.001; A1A2 vs. A1A2-H: p-value = < 0.001; A1A2-L vs. A1A2-H: p-value = 0.057).

Panel (c) of Table 4.2 presents the participation rate and average effort per minute conditional on participation in each treatment. In the absence of monetary incentives (treatment A1A2), the participation rate is 38%. The participation rate jumps to 98% since the impact of monetary incentives may be higher in an environment without these considerations. We thank a reviewer for pointing this out to us.
when low incentives are present (treatment A1A2-L) and 96% when high incentives are present (treatment A1A2-H). Hence, the participation rate increases to almost 100% in treatment A1A2-L, leaving no room for further improvement in treatment A1A2-H (Pairwise Fisher’s exact test, A1A2 vs. A1A2-L: p-value = < 0.001; A1A2 vs. A1A2-H: p-value = < 0.001; A1A2-L vs. A1A2-H: p-value = 0.608).

Conditional on participating, the average effort per minute is 0.54 in treatment A1A2. In the presence of monetary incentives, the average effort per minute increases to 2.52 in treatment A1A2-L and 3.27 in treatment A1A2-H. Pairwise Wilcoxon rank-sum tests reject the null hypotheses that the effort per minute between any pair of treatments are equal to each other (A1A2 vs. A1A2-L: p-value < 0.001; A1A2 vs. A1A2-H: p-value < 0.001; A1A2-L vs. A1A2-H: p-value = 0.020). Hence, on the intensive margin, we observe a monotonic response in subjects’ effort levels to changes in monetary incentives.
Table 4.5 shows that the results of OLS and quantile (median) regressions of participation (column 1) and effort per minute (columns 2 and 3) are in line with the results from the non-parametric tests. The table presents the test of equality between the estimated coefficients for treatments A1A2-L and A1A2-H. Treatment A1A2 is the reference treatment for all regressions.

In summary, Experiment 3 demonstrates that when non-monetary benefits are reduced and opportunity cost of effort is increased with the introduction of an alternative activity, we are able to detect a significant response to changes in monetary incentives.22 The introduction of monetary incentives in treatment A1A2-L increases participation to almost 100% and has a significant positive impact on effort per minute. Increasing monetary incentives from treatment A1A2-L to treatment A1A2-H has a further significant impact on effort per minute. These results stand in contrast to what we obtain in Experiment 1 with the baseline treatments (B, BL, and BH), where we fail to detect a consistent response to changes in monetary incentives.

4.6 Conclusion

In recent years, real-effort tasks have been growing in popularity. In this paper, we design three real-effort experiments spanning eight treatments to examine a fundamental question in economics about whether subjects respond to monetary incentives in their effort choice. Evidence from the laboratory has thus far provided mixed findings on the relationship between monetary incentives and effort provision. We re-visit this issue in Experiment 1 within a two-player real-effort contest and find that subjects exert significantly high effort even in the absence of monetary incentives, suggesting that they are motivated by non-monetary incentives. Moreover, we do not find any systematic changes in effort provision when monetary incentives are increased.

22Using a piece-rate scheme, DellaVigna and Pope (forthcoming) also find that subjects respond positively to increases in monetary incentives. This is consistent with our result in Experiment 3, but it contrasts the findings of Araujo et al. (2016) (who also use a piece rate) and our Experiment 1. A possible explanation is that their study was done using Amazon Mechanical Turk (MTurk) where subjects readily have outside options (e.g., to browse the Internet or perform other MTurk tasks/experiments), whereas the evidence presented in Araujo et al. (2016) is based on laboratory experiments where the subjects have a much lower opportunity cost of participating in the real-effort task (as in our Experiment 1).
Table 4.5: Regression results for Experiment 3

<table>
<thead>
<tr>
<th>Variables</th>
<th>Participation =1 if Effort &gt;0</th>
<th>Effort per minute given Effort &gt;0</th>
<th>Effort per minute given Effort &gt;0</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(OLS)</td>
<td>(OLS)</td>
<td>(Quantile)</td>
</tr>
<tr>
<td>Activity 1/2 – Low (A1A2-L)</td>
<td>0.641***</td>
<td>1.966***</td>
<td>1.934***</td>
</tr>
<tr>
<td></td>
<td>(0.063)</td>
<td>(0.423)</td>
<td>(0.664)</td>
</tr>
<tr>
<td>Activity 1/2 – High (A1A2-H)</td>
<td>0.603***</td>
<td>2.884***</td>
<td>3.280***</td>
</tr>
<tr>
<td></td>
<td>(0.064)</td>
<td>(0.417)</td>
<td>(0.655)</td>
</tr>
<tr>
<td>% invested in risk task</td>
<td>-0.126</td>
<td>1.291**</td>
<td>1.052</td>
</tr>
<tr>
<td></td>
<td>(0.106)</td>
<td>(0.592)</td>
<td>(0.929)</td>
</tr>
<tr>
<td>Age</td>
<td>0.015*</td>
<td>0.078</td>
<td>0.078</td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td>(0.049)</td>
<td>(0.077)</td>
</tr>
<tr>
<td>Female</td>
<td>-0.054</td>
<td>-0.163</td>
<td>-0.365</td>
</tr>
<tr>
<td></td>
<td>(0.056)</td>
<td>(0.309)</td>
<td>(0.485)</td>
</tr>
<tr>
<td>Economics major</td>
<td>-0.023</td>
<td>-0.405</td>
<td>-0.250</td>
</tr>
<tr>
<td></td>
<td>(0.083)</td>
<td>(0.455)</td>
<td>(0.715)</td>
</tr>
<tr>
<td>Undergraduate student</td>
<td>-0.015</td>
<td>-0.773</td>
<td>0.016</td>
</tr>
<tr>
<td></td>
<td>(0.229)</td>
<td>(1.093)</td>
<td>(1.717)</td>
</tr>
<tr>
<td>Graduate student</td>
<td>-0.007</td>
<td>-1.254</td>
<td>-0.533</td>
</tr>
<tr>
<td></td>
<td>(0.235)</td>
<td>(1.124)</td>
<td>(1.765)</td>
</tr>
<tr>
<td>Australian</td>
<td>-0.021</td>
<td>0.169</td>
<td>0.183</td>
</tr>
<tr>
<td></td>
<td>(0.056)</td>
<td>(0.302)</td>
<td>(0.474)</td>
</tr>
<tr>
<td># previous experiments</td>
<td>-0.007</td>
<td>-0.067</td>
<td>-0.059</td>
</tr>
<tr>
<td></td>
<td>(0.014)</td>
<td>(0.073)</td>
<td>(0.115)</td>
</tr>
<tr>
<td>Constant</td>
<td>0.164</td>
<td>-0.903</td>
<td>-1.632</td>
</tr>
<tr>
<td></td>
<td>(0.319)</td>
<td>(1.658)</td>
<td>(2.604)</td>
</tr>
<tr>
<td>Test:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A1A2-H = A1A2-L</td>
<td>-0.038</td>
<td>0.918***</td>
<td>1.345***</td>
</tr>
<tr>
<td></td>
<td>(0.062)</td>
<td>(0.303)</td>
<td>(0.475)</td>
</tr>
<tr>
<td>Observations</td>
<td>161</td>
<td>125</td>
<td>125</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.477</td>
<td>0.352</td>
<td></td>
</tr>
</tbody>
</table>

Ordinary least squares estimates given in (1) and (2), and quantile (median) regression estimates given in (3). Standard errors in parentheses. For all regressions, treatment B is the reference treatment.

In all regressions, the controls are treatment variables and subjects' characteristics, which include the percentage of the endowment invested in the risk task, age, gender, study level at University, whether the subject is pursuing a major in economics, whether the subject is Australian, and previous experience with economic experiments.

For treatment A1A2, we consider participation and effort per minute in Activity 1.

*** p<0.01, ** p<0.05, * p<0.10.
Interestingly, we find that monetary incentives have a positive impact on effort provision once the impact of non-monetary incentives is reduced. Experiments 2 and 3 reveal that introducing an alternative activity that pays a piece rate is the most effective way of mitigating the impact of non-monetary incentives by both increasing the opportunity cost of effort and directly decreasing the non-monetary benefits that are present. By altering both the benefit and cost side of effort provision, this treatment aims to replicate most modern work environments where individuals can choose how to spend their time at work. While paid outside options may not necessarily be available in workplaces, outside options can take other forms, such as chatting with co-workers, browsing the Internet, going for a coffee break etc. Although there may be no explicit monetary incentives associated with these outside options, individuals still derive some positive utility from them.

Workers in real-world workplaces are motivated by a myriad of factors (see, e.g., Besley and Ghatak, 2005; Akerlof and Kranton, 2010). Both monetary as well as non-monetary factors, such as whether an individual likes the job s/he is doing, play a role in effort provision. Nevertheless, when studying effort provision in the laboratory (using a real-effort task), we posit that the existence of non-monetary incentives may diminish the ability of researchers to detect any response to changes in monetary incentives. Hence, researchers may find it useful to study the impact of monetary and non-monetary incentives on effort provision separately. We provide a framework for doing this.

Beyond studying the economic question of the relationship between monetary incentives and effort provision, this paper also provides insights to researchers who are considering the use of real-effort tasks as part of their research strategy. Since subjects are motivated by non-monetary incentives in their real-effort choices, any impact this may have on outcomes should be taken into account. For instance, if the researcher is interested in using a real-effort task to assign subjects into different groups and/or roles, then s/he will need to be aware of potential selection effects since the sorting of subjects may be based on the subjects’ responses to non-monetary incentives. A failure to control or account for this selection effect can lead to selection based on unobservable
characteristics and consequently, false conclusions.

Consider, for example, a researcher who is interested in studying the impact of income earned on subsequent decisions (such as charitable giving). If non-monetary incentives, such as curiosity about the task, play a more important role in the determination of effort choices than monetary incentives, then the high earners may be those individuals who have a higher level of curiosity than the rest of the population. As a result, their decisions in the later stages of the experiment may be driven by this qualification rather than differences in income levels. If the researcher ignores this fact, then the causal effect may be misidentified. Mitigating the impact of non-monetary incentives in order to isolate the impact of monetary incentives can be useful in such cases.
Chapter 5

Conclusion

This thesis uses experimental methods to explore issues in leadership and contests within economics. Taken together, the three chapters provide exciting perspectives for future research. Chapter 2 contributes to the literature on leadership by examining attribution biases when leaders make decisions under risk on behalf of group members. We find an overall ‘bad news’ effect, in that members tend to attribute bad outcomes of a leader more to effort and good outcomes more to luck. Relative to the Bayesian benchmark, leaders are apportioned enough blame for their failures, but given too little credit for their successes. Understanding biases in the way leaders are attributed for their outcomes is important as these biases may shape the leader’s decision-making process. Further work is needed to examine under what circumstances would these attribution biases be welfare-reducing, i.e., by inducing leaders to make sub-optimal decisions.

The second and third chapters contribute to our understanding of the methodology issues pertaining to the optimal design of economic experiments. Chapter 3 evaluates and compares the performance of the quadratic and binarized scoring rules in eliciting subjects’ beliefs about the decisions made by others within a leadership environment. Overall, we find no statistically significant differences in the subjects’ behavior under the two scoring rules. However, our findings do not necessarily imply that the scoring rules do not affect the subjects’ behavior differently. We posit that there are several other factors that may affect the performance of the scoring rules within our environment, and investigating these issues is left for future work. Chapter 4 examines the response of subjects to
monetary incentives in real-effort tasks within a contest environment. We find evidence that subjects are motivated by non-monetary incentives in the laboratory, and that they respond to monetary incentives once the impact of the non-monetary incentives has been mitigated. Importantly, our study provides a useful framework for researchers who are interested in examining real-effort provision within a contest environment. This framework can also be applied to other environments, e.g., in leadership, where researchers may be interested in the factors motivating individuals to compete for positions of authority.
Bibliography


Appendices
Appendix A

Appendices for Chapter 2

A.1 Experimental instructions

This section contains the instructions used in the experiment described in Chapter 2.
Overview of Experiment

Thank you for agreeing to take part in this study which is funded by the Australian Research Council. Please read the following instructions carefully. A clear understanding of the instructions will help you make better decisions and increase your earnings from the experiment.

You will participate in two experiments today: Experiment 1 and Experiment 2. You will receive detailed instructions for each experiment before you participate in them. Note that your decisions in Experiment 2 will not change the earnings that you receive from Experiment 1. You will be informed of the outcomes of both experiments at the end of today’s session.

You will be paid for the decisions you make in either Experiment 1 or Experiment 2. This implies that you should carefully consider all of the decisions you make in both experiments as they may determine your earnings. Whether you will be paid for Experiment 1 or Experiment 2 will be randomly determined at the end of the session. Your final payment today will also include a $10 participation fee.

During the experiments, we will be using Experimental Currency Units (ECU). At the end of the session, we will convert the amount you earn into Australian Dollars (AUD) using the following conversion rate: 10 ECU = 1 AUD.

At the end of Experiment 2, you will be asked to fill out a brief questionnaire asking you some general questions. All of the decisions you make in today’s session will remain anonymous.

Please do not talk to one another during the experiment. If you have any questions, please raise your hand and we will come over to answer your questions privately.
Experiment 1

You will participate in Experiment 1 in groups of two. The computer will randomly match you with one other person in the room. You will never learn the identity of your partner.

Each of you is given an endowment of 300 ECU, and you are asked to divide this amount between yourself and the person you are matched with.

At the end of today’s session, if this experiment is picked for payment, then you will be paid either according to your decision or according to the decision made by your randomly matched partner. The computer will randomly determine whose allocation decision will be implemented.

**Example.** Suppose you choose to divide your endowment by keeping 200 ECU for yourself and giving 100 ECU to your matched partner. Your matched partner decides to keep 130 ECU and give 170 ECU to you. If, at the end of the experiment, the computer randomly determines that it is the allocation of your matched partner that gets implemented, then your payment will be 170 ECU and your matched partner’s payment will be 130 ECU.

Are there any questions? If not, we will proceed with Experiment 1.
Experiment 2

Experiment 2 consists of six identical rounds. At the end of the experiment, if you are paid for Experiment 2, then the computer will randomly pick one of the six rounds for payment.

You will participate in each round in groups of three. At the beginning of each round, the computer will randomly match you with two other people in this room with whom you have not been matched before. You will never learn the identity of your partners. Each round consists of three stages.

Stage 1: Appointment of a group leader.

In this stage, one group member will be assigned to be the leader of the group. There will be four possible methods to determine who is assigned the role of the leader. At the beginning of each round, the computer will reveal which method will be used to determine the leader for that round.

Method 1: One group member will be randomly assigned by the computer to be the leader. Hence, each group member has an equal chance of being assigned the role of the leader.

Method 2: The group member who transferred the lowest amount to his/her matched partner in Experiment 1 will be assigned to be the leader (ties will be broken randomly).

Method 3: The group member who transferred the highest amount to his/her matched partner in Experiment 1 will be assigned to be the leader (ties will be broken randomly).

Method 4: Each individual within the group will be asked to indicate whether you prefer your leader to be someone who has transferred the highest or the lowest amount to his/her matched partner in Experiment 1. The computer will then randomly pick one of the decisions of the group members to implement. If your decision is implemented, then one of your other two group members will be appointed to be the leader based on your preference. Hence, you will not be appointed to be the leader if your decision is implemented.
Example 1. Suppose the leader is appointed using Method 4. In Experiment 1, Player 1 chose to transfer 100 ECU to his/her matched partner, and Player 2 chose to transfer 160 ECU to his/her matched partner. Player 3 indicates that his/her preferred leader is someone who has transferred the lowest amount to his/her matched partner in Experiment 1. If the computer randomly determines that Player 3’s decision will be implemented, then Player 1 will be assigned the role of the leader.

You will only need to indicate your preferred leader for Method 4 once, at the beginning of Experiment 2. The same decision will be used whenever Method 4 is being used to determine the appointment of the group leader.

Stage 2: Investment decision by the group leader.

The leader will be given an endowment of 300 ECU. S/he will be asked to choose between two investment options that will affect the payoffs of all group members. Each investment can either fail or succeed. The two investment options have different chances of success/failure. They also have different costs to the leader.

Specifically, the two investments are:

**Investment X:** This investment costs 250 ECU to the leader. It will succeed with a 75% chance, and fail with a 25% chance.

**Investment Y:** This investment costs 50 ECU to the leader. It will succeed with a 25% chance, and fail with a 75% chance.

The payoffs to the leader and each group member in this stage of Experiment 2 are calculated as follows:

1. Payoff to leader = 300 ECU − Cost of investment + Returns on investment
2. Payoff to each group member = Returns on investment

Note that the amount that you receive from each investment may be different in each round, and this may affect the final payoffs to the leader and each group member. However, you will always receive a higher payoff if the investment succeeds, and a lower payoff if it fails. Please pay attention to these numbers on the screen in each round.

Figure A.1 shows an example where the returns of each investment options are 200 ECU if the investment succeeds, and 0 ECU if the investment fails, i.e., as shown by the numbers in red.
Example 2. Suppose in the round depicted in Figure A.1, the leader chooses Investment X for the group. Then, the investment costs the leader 250 ECU, and will succeed with a 75% chance and fail with a 25% chance. At the end of the experiment, if the investment succeeds, then each group member will receive 200 ECU, and the leader will receive 
\[(300 - 250 + 200) = 250\text{ ECU}\]
for this stage of Experiment 2.

Example 3. Suppose in the round depicted in Figure A.1, the leader chooses Investment Y for the group. Then, the investment costs the leader 50 ECU, and will succeed with a 25% chance and fail with a 75% chance. At the end of the experiment, if the investment fails, then each group member will receive 0 ECU, and the leader will receive 
\[(300 - 50 + 0) = 250\text{ ECU}\]
for this stage of Experiment 2.

You will be informed whether you have been assigned the role of the leader at the end of the experiment. Hence, you will be asked to make an investment decision in Stage 2 of each round assuming that you have been assigned the role of the leader. Your decision will be implemented if you have been assigned the role of the leader for that round.

At the end of the experiment, all group members will learn how much they have received from the chosen investment, but they will not learn the investment decision of the leader.
Stage 3: Beliefs of the other group members.

After you have made your investment decision, you will be asked to predict which investment your leader has chosen, assuming that someone else in your group has been assigned the role of the leader.

Specifically, we would like to know how likely it is in your opinion that the leader has chosen Investment X. Suppose there were 100 people in the position the leader is in now. How many of them do you think would choose Investment X?

You will need to choose a number between 0 and 100. A higher number means that you think the leader is more likely to have chosen Investment X.

The specific questions you will be asked are listed below.

Question 1
Suppose there were 100 people in the position the leader is in now. How many of them do you think would choose Investment X?

In Question 2, you are given additional information. You are asked to evaluate the same question with this additional information. Specifically, you should consider whether your guess of the leader’s decision will be different, given that you know the outcome of the investment chosen by your leader.

Question 2

Suppose you are informed that the investment chosen by your leader has succeeded, and you have therefore received the high payoff.

Now consider whether your guess will be higher than, lower than, or the same as the one you stated in Question 1. That is, suppose there were 100 people in the position the leader is in now. Given an outcome of high payoff, how many of them do you think have chosen Investment X?

Suppose you are informed that the investment chosen by your leader has failed, and you have therefore received the low payoff.

Now consider whether your guess will be higher than, lower than, or the same as the one you stated in Question 1. That is, suppose there were 100 people in the position the leader is in now. Given an outcome of low payoff, how many of them do you think have chosen Investment X?
The computer will randomly select one of these two questions and you will be paid for your response to this question. If Question 2 is chosen for payment, then you will be paid for your answer to the scenario that corresponds to the actual outcome of the investment chosen by your leader.

The section below describes how your payoff in Stage 3 will be determined. This procedure has been used in many other studies. We explain the procedure in detail, but what is most important is that this payoff structure is designed such that it is in your best interest to report your true belief about your leader’s decision.

Your payment for the question randomly chosen by the computer is determined as follows. You will receive 10 ECU with some chance. Your chance of receiving 10 ECU depends on your answer and the leader’s decision. The closer your guess is to the actual decision made by your leader, the higher is your chance of receiving the fixed payment of 10 ECU.

Specifically, your chance of receiving 10 ECU is determined by the following formula:

\[
\text{Chance of receiving 10 ECU} = \left[ 1 - \left( \frac{x - \text{your guess}}{100} \right)^2 \right] \times 100.
\]

\(x\) takes the value of 100 if your leader chose Investment X, and \(x\) takes the value of 0 if your leader chose Investment Y.

To illustrate, suppose your leader has chosen Investment X. This means that \(x = 100\) in the formula above, and your chance of receiving 10 ECU will be higher if your guess is higher. If you state 100 as your guess that the leader has chosen Investment X, then your chance of receiving 10 ECU will be \([1 - \left( \frac{100-100}{100} \right)^2] \times 100 = 100\). On the other hand, suppose your leader has chosen Investment Y instead, while your guess remains at 100. This means that \(x = 0\) in the formula above, and your chance of receiving 10 ECU will be \([1 - \left( \frac{0-100}{100} \right)^2] \times 100 = 0\).

Here is another example:

**Example 4.**  Suppose you guess 70 as the chance that your leader has chosen Investment X for the group. At the end of the experiment, the computer reveals that your leader has chosen Investment X for the group. Hence, your chance of receiving 10 ECU will be \([1 - \left( \frac{100-70}{100} \right)^2] \times 100 = 91\).
To determine whether you receive 10 ECU, the computer will randomly draw a number between 0 and 100 (including decimal points). If the number drawn by the computer is less than or equal to your chance of receiving 10 ECU as determined by the formula above, then you will receive 10 ECU. Otherwise, you will receive 0 ECU. Hence, in Example 4 above, if the number randomly drawn by the computer is less than or equal to 91, then you will receive 10 ECU. Otherwise, you will receive 0 ECU.

**Payment for Experiment 2:**

At the end of the experiment, if you are paid for Experiment 2, then the computer will randomly select one of the six rounds for payment. For the randomly chosen round:

1. If you are assigned the role of the leader, then you will be paid according to your investment decision in Stage 2 only.

2. If you are not assigned the role of the leader, then you will be paid according to your leader’s investment decision in Stage 2, plus your decisions in Stage 3. The computer will randomly select one of the two questions in Stage 3, and you will be paid for your response to this question.
Summary

1. You will participate in six identical rounds in Experiment 2. At the beginning of each round, the computer will randomly match you to a new group with two other people. Each round consists of three stages.

2. In Stage 1, one group member will be assigned to be the leader of the group. There are four possible methods to determine who is assigned the role of the leader. You will be informed which method will be used to determine the leader at the beginning of each round.

   In Method 1, the computer will randomly assign one group member to be the leader.

   In Method 2, the group member who transferred the lowest amount to his/her matched partner in Experiment 1 will be assigned to be the leader.

   In Method 3, the group member who transferred the highest amount to his/her matched partner in Experiment 1 will be assigned to be the leader.

   In Method 4, you will be asked to indicate whether you prefer your leader to be someone who has transferred the highest or the lowest amount to his/her matched partner in Experiment 1. The computer will pick one of the decisions of the group members to implement. If your decision is implemented, then one of your other two group members will be appointed to be the leader based on your preference. Hence, you will not be appointed to be the leader if your decision is implemented.

   You will be asked to indicate your preferred leader for Method 4 once, at the beginning of Experiment 2. The computer will use the same decision whenever Method 4 is being used to determine the leader.

3. In Stage 2, you will be asked to make an investment decision, assuming that you have been assigned the role of the leader. The leader will be given an endowment of 300 ECU, and s/he will be asked to choose between two investment options that will affect the payoffs of all group members. Your decision will be implemented for your group only if you have been assigned the role of the leader for that round.

4. Investment X and Investment Y may be different in each round. In each round, the amount that you receive from each investment may be different, but you will always receive a higher payoff if the investment succeeds, and a lower payoff if it fails. The investment options will be shown on your computer screens.
5. In Stage 3, you will be asked to predict which investment your leader has chosen, assuming that you have not been assigned the role of the leader. You will be asked two questions.

In Question 1, you will be asked to predict how likely it is in your opinion that the leader has chosen Investment X. You will need to choose a number between 0 and 100. A higher number means that you think the leader is more likely to have chosen Investment X.

In Question 2, you are given additional information. Specifically, you will be asked the same question under two different scenarios: (i) suppose you are told that the investment has succeeded; and (ii) suppose you are told that the investment has failed. You should consider whether your guess of the leader’s decision will be higher than, lower than, or the same as the one you stated in Question 1, given that you know the outcome of the investment chosen by your leader.

6. The payoff structure used to determine your payment in Stage 3 is designed such that it is in your best interest to report your true beliefs about your leader’s decision.

7. At the end of the experiment, the computer will randomly select one of the six rounds for payment. For the randomly chosen round, if you are assigned the role of the leader, then you will be paid according to your decision in Stage 2. If you are not assigned the role of the leader, then you will be paid according to your leader’s decision in Stage 2, as well as your decisions in Stage 3. The computer will randomly select one of the two questions in Stage 3 for payment.

If you have any questions, please raise your hand and an experimenter will come to you to answer your questions privately. Otherwise, please wait patiently for the experimenter to launch the practice questions on your computer screens. The purpose of these practice questions is to make sure that you understand the experiment. If you have any questions at any time, please raise your hand and an experimenter will come over to answer your questions privately.

Once everyone has completed the practice questions, we will proceed with one practice round for Experiment 2. The purpose of the practice round is to allow you to familiarize yourself with the decision screens. Your decisions in the practice round will not affect your payments for today’s experiment. We will proceed with Experiment 2 once everyone has completed the practice round.
Practice Questions (Experiment 2)

1. I will be paid for the decisions in both experiments today. True/False [Ans: False]

2. We will participate in six identical rounds in Experiment 2. If we are paid for Experiment 2, then we will be paid for our decisions in one of the six rounds. True/False. [Ans: True]

3. We will participate in each round of Experiment 2 in groups of three. One group member will be assigned the role of the leader. True/False [Ans: True]

4. In Experiment 1, Player 1 chose to transfer 160 ECU to his/her matched partner, Player 2 chose to transfer 115 ECU to his/her matched partner, and Player 3 chose to transfer 160 ECU to his/her matched partner.

Suppose the leader is appointed using Method 2. Which of the following is correct? [Ans: (b)]

(a) Player 1 will be assigned the role of the leader.
(b) Player 2 will be assigned the role of the leader.
(c) Player 3 will be assigned the role of the leader.
(d) Both Player 1 and Player 3 have an equal chance of being assigned the role of the leader.

5. In the above example, suppose the leader is appointed using Method 3. Which of the following is correct? [Ans: (d)]

(a) Player 1 will be assigned the role of the leader.
(b) Player 2 will be assigned the role of the leader.
(c) Player 3 will be assigned the role of the leader.
(d) Both Player 1 and Player 3 have an equal chance of being assigned the role of the leader.

6. Suppose the leader is appointed using Method 4. Suppose also that your preference for leadership appointment is randomly chosen by the computer to be implemented. Which of the following is correct? [Ans: (b)]

(a) Depending on what I indicate as my preference of the appointed leader, I have a chance of being assigned the role of the leader.
(b) Regardless of what I indicate as my preference of the appointed leader, I will definitely not be assigned the role of the leader.
7. Suppose the leader is appointed using Method 4. In Experiment 1, Player 1 chose to transfer 200 ECU to his/her matched partner, and Player 2 chose to transfer 85 ECU to his/her matched partner. Player 3 indicates that his/her preferred leader is someone who has transferred the highest amount to his/her matched partner in Experiment 1.

Suppose Player 3’s decision is randomly chosen by the computer to be implemented. Which of the following is correct? [Ans: (a)]

(a) Player 1 will be assigned the role of the leader.
(b) Player 2 will be assigned the role of the leader.
(c) Both Player 1 and Player 2 have an equal chance of being assigned the role of the leader.

8. Which of the following is correct? [Ans: (b)]

(a) The other group members will be informed of the investment chosen by the leader, but not the amount they have received from the investment.
(b) The other group members will be informed of the amount they have received from the investment chosen by the leader, but not the investment chosen by him/her.
(c) The other group members will be informed of the investment chosen by the leader, and the amount they have received from the investment.
9. Consider the investment options depicted in the figure below.

![Investment Options Diagram](image)

**Figure A.2: Investment Options (Practice Question)**

Suppose the leader chooses Investment X.

(a) At the end of the experiment, the computer randomly determines that the investment succeeds.

If you are not the leader, how many ECU will you receive from Stage 2 of Experiment 2? [Ans: 250 ECU]

(b) At the end of the experiment, the computer randomly determines that the investment fails.

If you are the leader, how many ECU will you receive from Stage 2 of Experiment 2? [Ans: 100 ECU]
10. Which of the following is true? [Ans: (c)]

(a) I will be paid for my decision in Stage 3 of Experiment 2 regardless of whether I have been assigned the role of the leader or not.
(b) I will be paid for my decision in Stage 3 of Experiment 2 only if I have been assigned the role of the leader.
(c) I will be paid for my decision in Stage 3 of Experiment 2 only if I have not been assigned the role of the leader.

11. In Stage 3, I will be asked two questions. If I am paid for Stage 3 of Experiment 2, then I will be paid according to my answers to both questions. True/False [Ans: False]

12. Suppose you strongly believe that the leader of your group has chosen Investment Y. Which of the following statement is true? [Ans: (b)]

(a) It is in my best interest to choose a higher number as my guess of “how likely is my leader to have chosen Investment X”.
(b) It is in my best interest to choose a lower number as my guess of “how likely is my leader to have chosen Investment X”.
(c) It is in my best interest to choose 50 as my guess of “how likely is my leader to have chosen Investment X”.
A.2 Screenshots for belief elicitation task

(a) Interim Belief

(b) Posterior Beliefs

Figure A.3: Decision screens – Elicitation of beliefs
A.3 Derivation of Hypothesis 2.1

We are interested in how members form interim beliefs about their leader’s type under each appointment mechanism, given that they know someone else in the group has been appointed to be the leader. Specifically, for any given appointment mechanism $\Psi$, we are interested in member $i$’s interim belief that their leader is of type $\beta \geq \beta^*$, and we denote this belief as $\mu^\Psi_i$.

**Random appointment (RA).** Under this mechanism, each individual has an equal chance of being appointed as the leader. Hence, member $i$’s interim belief is simply their belief about the type of any given individual. This implies that

$$\mu^\text{RA}_i = \Pr(\beta \geq \beta^*) = 1 - F(\beta^*). \quad (\text{A.1})$$

For all other appointment mechanisms, we will compare the interim beliefs to that of the baseline treatment RA.

**Appointment of lowest type (LA).** Under this mechanism, the individual with the lowest $\beta$ in the group is appointed to be the leader.

Consider member $i$ of type $\beta_i$ who has been informed that someone else in the group is appointed to be the leader. Member $i$ therefore knows that the leader has a type $\beta \leq \beta_i$, as otherwise they would have been appointed to be the leader instead.

Given this, there are two possible cases. First, if $\beta_i < \beta^*$, then it must be that $\mu^\text{LA}_i = 0$ since the leader has a type $\beta \leq \beta_i < \beta^*$. Second, if $\beta_i \geq \beta^*$, then the probability that the appointed leader is of type $\beta \geq \beta^*$ depends on the probability of drawing a type $\beta \geq \beta^*$ from the truncated support $[0, \beta_i]$, since member $i$ knows that the leader has a type $\beta \leq \beta_i$. Using the probability distribution on a truncated support, this probability is given by $1 - \frac{F(\beta^*)}{F(\beta_i)}$. 
Hence, for member $i$ of type $\beta_i$,

$$
\mu_{i}^{LA} = \begin{cases} 
0 & \text{if } \beta_i < \beta^*, \\
1 - \frac{F(\beta^*)}{F(\beta_i)} & \text{if } \beta_i \geq \beta^*.
\end{cases}
$$

(A.2)

We can easily observe that $\mu_{i}^{LA} < \mu_{i}^{RA}$ for $\beta_i < \beta^*$, with equality only if $\beta^* = 1$. For $\beta_i \geq \beta^*$, note that $\frac{F(\beta^*)}{F(\beta_i)} > F(\beta^*)$ with equality only if $\beta_i = 1$. This implies that $1 - \frac{F(\beta^*)}{F(\beta_i)} < 1 - F(\beta^*)$. Hence, $\mu_{i}^{LA} < \mu_{i}^{RA}$ for $\beta_i \geq \beta^*$.

**Appointment of highest type (HA).** Under this mechanism, the individual with the highest $\beta$ in the group is appointed to be the leader.

Consider member $i$ of type $\beta_i$ who has been informed that someone else in the group is appointed to be the leader. Member $i$ therefore knows that the leader has a type $\beta \geq \beta_i$, as otherwise they would have been appointed to be the leader instead.

Given this, there are two possible cases. First, if $\beta_i \geq \beta^*$, then it must be that $\mu_{i}^{HA} = 1$ since the leader has a type $\beta \geq \beta_i \geq \beta^*$. Second, if $\beta_i < \beta^*$, then the probability that the appointed leader is of type $\beta \geq \beta^*$ depends on the probability of drawing a type $\beta \geq \beta^*$ from the truncated support $[\beta_i, 1]$, since member $i$ knows that the leader has a type $\beta \geq \beta_i$. Using the probability distribution on a truncated support, this probability is given by $1 - \frac{F(\beta^*) - F(\beta_i)}{1 - F(\beta_i)} = \frac{1 - F(\beta^*)}{1 - F(\beta_i)}$.

Hence, we have for member $i$ of type $\beta_i$,

$$
\mu_{i}^{HA} = \begin{cases} 
1 & \text{if } \beta_i \geq \beta^*, \\
\frac{1 - F(\beta^*)}{1 - F(\beta_i)} & \text{if } \beta_i < \beta^*.
\end{cases}
$$

(A.3)

We can easily observe that $\mu_{i}^{HA} > \mu_{i}^{RA}$ for $\beta_i \geq \beta^*$, with equality only if $\beta^* = 0$. For $\beta_i < \beta^*$,

$$
\mu_{i}^{HA} - \mu_{i}^{RA} = \frac{1 - F(\beta^*)}{1 - F(\beta_i)} - [1 - F(\beta^*)] = [1 - F(\beta^*)]\left[\frac{1}{1 - F(\beta_i)} - 1\right] > 0,
$$
since $\frac{1}{1-F(\beta)} > 1$, with equality only if $\beta = 0$. Hence, $\mu_i^{HA} > \mu_i^{RA}$ for $\beta_i < \beta^*$. 

**Group appointment (GA).** Under this mechanism, all individuals indicate how they want their leader to be appointed. Specifically, they may nominate either: (i) the lowest type as the leader; (ii) the highest type as the leader; or (iii) indifference. One of the group members’ nomination decision is randomly chosen to be implemented and the leader is appointed from the remaining group members based on this individual’s preference.

Regardless of their type, it is trivial to see that all individuals will prefer to have the highest type appointed as the leader. Intuitively, this is because all group members will want a leader of type $\beta \geq \beta^*$ such that the leader will choose a higher effort level $e_H$ over $e_L$, leading to higher expected utility for the members.

Given this, now consider member $i$ of type $\beta_i$ who has been informed that someone else in the group is appointed to be the leader under this appointment mechanism. Member $i$ updates their belief that their own nomination decision has been randomly implemented given this information. Assuming Bayesian updating, their posterior belief that their nomination decision has been implemented is given by $\frac{1}{1+(N-1)[1-F(N-2,\beta_i)]}$. To see this, first consider the prior probability that member $i$’s decision was implemented, which is given by $\frac{1}{N}$. A priori, if individual $i$’s decision is implemented, then they become a member (non-leader) with probability 1. On the other hand, another individual’s nomination decision is implemented with a prior probability given by $\frac{N-1}{N}$. A priori, if another individual’s nomination decision is implemented, then individual $i$ becomes a member with probability $1 - F^{N-2}(\beta_i)$, i.e., this is the probability that at least someone else other than both the nominator and individual $i$ has a type $\geq \beta_i$. Hence, conditional on member $i$ knowing that they are not appointed to be the leader, the posterior probability that their nomination decision is implemented can be derived using Bayes’ rule, i.e.,

$$\frac{\frac{1}{N}(1)}{\frac{1}{N}(1) + \frac{(N-1)}{N}[1-F^{N-2}(\beta_i)]}.$$ 

There are now two possible cases. First, if individual $i$’s nomination decision is implemented, then the subjective probability that the appointed leader is of type $\beta \geq \beta^*$ depends on the probability that at least one of the other $N-1$ group members is of type $\geq \beta^*$, and this is given by $1 - F^{N-1}(\beta^*)$. Second, if individual $i$’s nomination decision is
not implemented, then the member \( i \) knows that the leader is of type \( \beta \geq \beta_i \), as otherwise they would have been appointed to be the leader instead (since all individuals would have chosen to appoint the highest type as the leader). Consequently, the probability that the appointed leader is of type \( \beta \geq \beta^* \) depends on the individual \( i \)'s own type, and the derivation follows that of mechanism HA. In other words, this probability is equals to 1 if \( \beta_i \geq \beta^* \), and equals to \( \frac{1-F(\beta^*)}{1-F(\beta^*)} \) if \( \beta_i < \beta^* \).

Putting all these together, we have for member \( i \) of type \( \beta_i \),

\[
\mu_i^{GA} = A \times \left[ 1 - F^{N-1}(\beta^*) \right] + (1 - A) \times \begin{cases} 
1 & \text{if } \beta_i \geq \beta^*, \\
\frac{1-F(\beta^*)}{1-F(\beta_i)} & \text{if } \beta_i < \beta^*, 
\end{cases}
\]  

(A.4)

where \( A \equiv \frac{1}{1+(N-1)[1-F^{N-2}(\beta_i)]} \).

First, note that \( F^{N-1}(\beta^*) < F(\beta^*) \) for all \( N > 2 \), with equality only if \( \beta^* = 0 \) or \( \beta^* = 1 \). This implies that \( 1 - F^{N-1}(\beta^*) > 1 - F(\beta^*) \). The last term in (A.4) is equivalent to (A.3), which we have already shown to be greater than \( 1 - F(\beta^*) \) for both cases. Hence, we observe that \( \mu_i^{GA} > \mu_i^{RA} \).
A.4 Additional analysis

A.4.1 Analysis with Game 1 treatments only

This section presents the analyses including only the Game 1 treatments. Table A.1 presents marginal-effects estimates from a probit model for the relationship between the subjects’ dictator game behavior and their decisions as leaders in the leadership task. Column (1) presents the regression at the pooled level, while columns (2) to (5) present the estimates for each appointment mechanism. The estimates reveal that, at the pooled level, leaders are 0.4% more likely to choose $e_H$ for every 1% of their endowment transferred in the dictator game, and this effect is statistically significant (p-value < 0.001). This finding is also robust at the treatment level in columns (2) to (5), where leaders are between 0.3% and 0.6% significantly more likely to choose $e_H$ on average for every additional 1% of their endowment transferred in the dictator game.

Table A.1: Regression of leader’s effort choice (Game 1)

<table>
<thead>
<tr>
<th>Variables</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>% endowment transferred in DG</td>
<td>0.004***</td>
<td>0.005***</td>
<td>0.006***</td>
<td>0.003**</td>
<td>0.004***</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(4.15)</td>
<td>(4.41)</td>
<td>(1.97)</td>
<td>(2.91)</td>
</tr>
<tr>
<td>% endowment invested in RG</td>
<td>-0.001</td>
<td>-0.002**</td>
<td>0.000</td>
<td>0.000</td>
<td>-0.001</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(2.40)</td>
<td>(0.15)</td>
<td>(0.36)</td>
<td>(0.90)</td>
</tr>
<tr>
<td>Order Effects</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Observations</td>
<td>1,088</td>
<td>272</td>
<td>272</td>
<td>272</td>
<td>272</td>
</tr>
</tbody>
</table>

Marginal effects of probit model reported. Robust standard errors in parentheses. Standard errors are clustered at the subject level in column (1).

DG: Dictator Game; RG: Risk Game.

*** p<0.01, ** p<0.05, * p<0.10.

Table A.2 presents OLS estimates for the regressions of interim beliefs against treatment variables. Similar to the main analysis in the paper, columns (1) and (3) control for order effects of the treatments while columns (2) and (4) control for individual fixed effects. In all the specifications, treatment RA is the comparison group in all the specifications. Moreover, we control for the leadership game that is implemented. The estimates reveal that Result 2.1 in the paper is robust to the exclusion of the Game 0 treatments.
Table A.2: Regression of members’ interim belief (Game 1)

<table>
<thead>
<tr>
<th>Variables</th>
<th>Dependent variable: Interim belief</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td></td>
<td>(1.484)</td>
</tr>
<tr>
<td></td>
<td>(1.397)</td>
</tr>
<tr>
<td>Treatment GA</td>
<td>2.717**</td>
</tr>
<tr>
<td></td>
<td>(1.355)</td>
</tr>
<tr>
<td>Chooses high effort as leader</td>
<td>24.584***</td>
</tr>
<tr>
<td></td>
<td>(1.960)</td>
</tr>
<tr>
<td>% endowment invested in RG</td>
<td>-0.086*</td>
</tr>
<tr>
<td></td>
<td>(0.045)</td>
</tr>
<tr>
<td>Constant</td>
<td>55.525***</td>
</tr>
<tr>
<td></td>
<td>(3.990)</td>
</tr>
<tr>
<td>Order Effects</td>
<td>Y</td>
</tr>
<tr>
<td>Individual FE</td>
<td>N</td>
</tr>
<tr>
<td>Observations</td>
<td>1,088</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.111</td>
</tr>
</tbody>
</table>

Robust standard errors clustered at the subject level in parentheses. For all regressions, treatment RA is the reference treatment.
RG: Risk Game.
*** p<0.01, ** p<0.05, * p<0.10.

Table A.3 presents the regression results of equation (2.7) with data from the Game 1 treatments. Note here that, as with the main analysis in the paper, the inconsistent and non-updaters have been dropped from this analysis. The estimates suggest that the exclusion of the Game 0 treatments does not change our conclusions in the paper and that Result 2.2 still holds even with the inclusion of these treatments. Hence, members suffer from base-rate neglect ($\delta < 1$: p-value < 0.001), attribute good outcomes more to the leader’s luck than a Bayesian ($\gamma_G < 1$: p-value < 0.001) but bad outcomes like a Bayesian ($\gamma_B = 1$: p-value = 0.492). Consequently, they tend to attribute good (bad) outcomes more to the leader’s luck (effort), i.e., $\gamma_G < \gamma_B$ (p-value = 0.012).

Moreover, members suffer from similar biases at the treatment level across all four appointment mechanisms. However, unlike Result 2.3, members do not suffer from any asymmetric bias in the attribution of good and bad outcomes in treatment LA, although
this effect is only marginally statistically insignificant (p-value = 0.103). In addition, members tend to attribute good (bad) outcomes more to luck (effort) in treatment HA, and this effect is now statistically significant (p-value = 0.028).

Table A.3: Regression of members’ posterior beliefs (Game 1)

<table>
<thead>
<tr>
<th>Variables</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \delta : ) Logit(interim belief)</td>
<td>0.733***</td>
<td>0.764***</td>
<td>0.793***</td>
<td>0.771**</td>
<td>0.529***</td>
</tr>
<tr>
<td></td>
<td>(0.049)</td>
<td>(0.071)</td>
<td>(0.057)</td>
<td>(0.093)</td>
<td>(0.135)</td>
</tr>
<tr>
<td>( \gamma_G : ) Good outcome ( \times ) logit(( p ))</td>
<td>0.742***</td>
<td>0.744***</td>
<td>0.728***</td>
<td>0.752***</td>
<td>0.798**</td>
</tr>
<tr>
<td></td>
<td>(0.060)</td>
<td>(0.089)</td>
<td>(0.078)</td>
<td>(0.094)</td>
<td>(0.098)</td>
</tr>
<tr>
<td>( \gamma_B : ) Bad outcome ( \times ) logit(1 - ( p ))</td>
<td>0.948</td>
<td>0.932</td>
<td>0.937</td>
<td>0.994</td>
<td>0.876</td>
</tr>
<tr>
<td></td>
<td>(0.076)</td>
<td>(0.092)</td>
<td>(0.119)</td>
<td>(0.090)</td>
<td>(0.114)</td>
</tr>
<tr>
<td>Observations</td>
<td>1,640</td>
<td>410</td>
<td>410</td>
<td>410</td>
<td>410</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.636</td>
<td>0.686</td>
<td>0.741</td>
<td>0.613</td>
<td>0.421</td>
</tr>
</tbody>
</table>

Test of \( \gamma_G = \gamma_B \)

| test statistic | -2.522 | -1.588 | -1.637 | -2.218 | -0.512 |
| p-value | 0.012 | 0.114 | 0.103 | 0.028 | 0.609 |

Robust standard errors clustered at the subject level in parentheses.
This analysis includes only the Game 1 treatments but includes subjects classified as inconsistent or non-updaters.
p<0.01, ** p<0.05, * p<0.10. Null hypothesis is coefficient = 1.

Finally, Table A.4 presents the results of a two-component finite mixture model analysis using only the Game 1 treatments. Even with the exclusion of the Game 0 treatments, the finite mixture model analysis reveals two groups of updates that are largely similar to that in Table 2.5. The majority of the belief updates in the sample is characterized by a modest level of base-rate neglect, under-responsiveness to the leader’s outcomes, as well as an asymmetric attribution of the leader’s outcomes to effort and luck. A small proportion of updates is characterized by a high level of base-rate neglect, over-responsiveness to the leader’s outcomes, and no asymmetric attribution of the leader’s outcomes to effort and luck.
Table A.4: Finite mixture model for updating behavior (Game 1)

<table>
<thead>
<tr>
<th>Variables</th>
<th>Component 1</th>
<th>Component 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\delta$: Logit (interim)</td>
<td>0.946***</td>
<td>0.208***</td>
</tr>
<tr>
<td></td>
<td>(0.011)</td>
<td>(0.120)</td>
</tr>
<tr>
<td>$\gamma_G$: Good outcome $\times \logit(p)$</td>
<td>0.528***</td>
<td>2.099**</td>
</tr>
<tr>
<td></td>
<td>(0.033)</td>
<td>(0.517)</td>
</tr>
<tr>
<td>$\gamma_B$: Bad outcome $\times \logit(1-p)$</td>
<td>0.651***</td>
<td>1.983**</td>
</tr>
<tr>
<td></td>
<td>(0.051)</td>
<td>(0.455)</td>
</tr>
<tr>
<td>Test of $\gamma_G = \gamma_B$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>test statistic</td>
<td>-2.770</td>
<td>0.230</td>
</tr>
<tr>
<td>p-value</td>
<td>0.006</td>
<td>0.819</td>
</tr>
</tbody>
</table>

Latent Class Marginal Probabilities

| $\mu$                                           | 0.884               | 0.116               |
|                                                 | (0.024)             | (0.024)             |

Model Fit

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Log likelihood</td>
<td>-2188.729</td>
<td></td>
</tr>
<tr>
<td>AIC</td>
<td>4395.457</td>
<td></td>
</tr>
<tr>
<td>BIC</td>
<td>4444.079</td>
<td></td>
</tr>
</tbody>
</table>

Robust standard errors clustered at the subject level in parentheses. This analysis excludes subjects classified as inconsistent or non-updaters.

*** p<0.01, ** p<0.05, * p<0.10. Null hypothesis is coefficient = 1.

A.4.2 Analysis of posterior beliefs with entire sample (including inconsistent and non-updaters)

Table A.5 presents the regression results of equation (2.7) with the inconsistent and non-updaters included. The estimates suggest that the inclusion of these subjects leads to an attenuation of the coefficient estimates for $\gamma_G$ and $\gamma_B$. Consequently, the coefficient estimates in column (1) now reveals that, at the pooled level, members tend to attribute a leader’s failure more to luck as compared to a Bayesian ($\gamma_B < 1$: p-value = 0.026). At the treatment level (columns 2 to 5), we observe that this bias is also present in treatments RA, HA, and GA (p-values = 0.056, 0.087, and 0.011, respectively). In treatment HA, the bias is in the similar direction although it is statistically insignificant (p-value = 0.379).

Finally, despite the attenuation in the coefficient estimates for $\gamma_G$ and $\gamma_B$, we find
Table A.5: Regression of members’ posterior beliefs (entire sample)

<table>
<thead>
<tr>
<th>Variables</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pooled</td>
<td>RA</td>
<td>LA</td>
<td>HA</td>
<td>GA</td>
</tr>
<tr>
<td>( \delta ): Logit(interim belief)</td>
<td>0.701***</td>
<td>0.737***</td>
<td>0.709***</td>
<td>0.716***</td>
<td>0.539***</td>
</tr>
<tr>
<td></td>
<td>(0.039)</td>
<td>(0.068)</td>
<td>(0.047)</td>
<td>(0.060)</td>
<td>(0.106)</td>
</tr>
<tr>
<td>( \gamma_G ): Good outcome ( \times ) logit(( p ))</td>
<td>0.530***</td>
<td>0.548***</td>
<td>0.358***</td>
<td>0.618***</td>
<td>0.662***</td>
</tr>
<tr>
<td></td>
<td>(0.064)</td>
<td>(0.086)</td>
<td>(0.094)</td>
<td>(0.083)</td>
<td>(0.093)</td>
</tr>
<tr>
<td>( \gamma_B ): Bad outcome ( \times ) logit(1 ( - ) ( p ))</td>
<td>0.848** 0.830*</td>
<td>0.903</td>
<td>0.867*</td>
<td>0.742**</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.068)</td>
<td>(0.089)</td>
<td>(0.110)</td>
<td>(0.078)</td>
<td>(0.100)</td>
</tr>
</tbody>
</table>

Observations 3,264 544 1,088 1,088 544
R-squared 0.550 0.620 0.606 0.488 0.382

Test of \( \gamma_G = \gamma_B \)

test statistic -3.218 -2.132 -3.376 -2.060 -0.550
p-value 0.001 0.034 0.001 0.040 0.583

Robust standard errors clustered at the subject level in parentheses. This analysis includes all subjects.

\( p < 0.01, ** p < 0.05, * p < 0.10 \). Null hypothesis is coefficient = 1.

that members attribute the outcomes of the leader asymmetrically in that they tend to attribute good (bad) outcomes more to luck (effort). As in the main paper, this effect is statistically significantly at the pooled level (p-value = 0.001) and in treatment LA (p-value = 0.001). However, we now also find this effect in treatments RA and HA (p-values = 0.034 and 0.040, respectively).

A.4.3 IV regression of posterior beliefs

Table A.6 presents the IV estimates of (2.7), where we use the leadership appointment mechanisms as instruments for the logit of members’ interim beliefs. We find that the conclusions from the IV estimates are similar to those obtained from the OLS estimates in column (1) of Table 2.4. There is statistical significant evidence that members suffer from base-rate neglect (\( \delta < 1: p\text{-value} < 0.001 \)). Moreover, they attribute good outcomes more to luck than a Bayesian would (\( \gamma_G < 1: p\text{-value} < 0.001 \)), but are no different from a Bayesian in attributing bad outcomes to effort and luck (\( \gamma_B = 1: p\text{-value} = 0.267 \)). Finally, we find statistically significant evidence that members tend to attribute a leader’s good outcomes more to luck and bad outcomes more to effort (\( \gamma_G < \gamma_B: p\text{-value} = 0.042 \)).
Table A.6: IV regression of members’ posterior beliefs

<table>
<thead>
<tr>
<th>Variables</th>
<th>Pooled</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\delta$ : Logit(interim belief)</td>
<td>0.792*** (0.046)</td>
</tr>
<tr>
<td>$\gamma_G$ : Good outcome $\times$ logit($p$)</td>
<td>0.787*** (0.056)</td>
</tr>
<tr>
<td>$\gamma_B$ : Bad outcome $\times$ logit(1 $-$ $p$)</td>
<td>0.929 (0.064)</td>
</tr>
</tbody>
</table>

Observations 2,460

Test of $\gamma_G = \gamma_B$

<table>
<thead>
<tr>
<th>test statistic</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>-2.030</td>
<td>0.042</td>
</tr>
</tbody>
</table>

Robust standard errors clustered at the subject level in parentheses. This analysis excludes subjects classified as inconsistent or non-updaters.

$*$ $p<0.10$, $**$ $p<0.05$, $***$ $p<0.01$. Null hypothesis is coefficient = 1.
Appendix B

Appendices for Chapter 3

B.1 Experimental instructions

The reader may refer to Appendix A.1 of Chapter 2 for the full experimental instructions used in treatment BSR. Here, we provide detailed instructions for Stage 3 of the leadership task, which differ between treatments QSR and BSR.
Experiment 2 (Stage 3) – Treatment QSR

After you have made your investment decision, you will be asked to predict which investment your leader has chosen, assuming that someone else in your group has been assigned the role of the leader.

Specifically, we would like to know how likely it is in your opinion that the leader has chosen Investment X. *Suppose there were 100 people in the position the leader is in now. How many of them do you think would choose Investment X?*

You will need to choose a number between 0 and 100. A higher number means that you think the leader is more likely to have chosen Investment X.

The specific questions you will be asked are listed below.

**Question 1**

Suppose there were 100 people in the position the leader is in now. How many of them do you think would choose Investment X?

In Question 2, you are given additional information. You are asked to evaluate the same question with this additional information. Specifically, you should consider whether your guess of the leader’s decision will be different, given that you know the outcome of the investment chosen by your leader.

**Question 2**

*Suppose you are informed that the investment chosen by your leader has succeeded, and you have therefore received the high payoff.*

Now consider whether your guess will be higher than, lower than, or the same as the one you stated in Question 1. That is, suppose there were 100 people in the position the leader is in now. Given an outcome of high payoff, how many of them do you think have chosen Investment X?

*Suppose you are informed that the investment chosen by your leader has failed, and you have therefore received the low payoff.*

Now consider whether your guess will be higher than, lower than, or the same as the one you stated in Question 1. That is, suppose there were 100 people in the position the leader is in now. Given an outcome of low payoff, how many of them do you think have chosen Investment X?
The computer will randomly select one of these two questions and you will be paid for your response to this question. If Question 2 is chosen for payment, then you will be paid for your answer to the scenario that corresponds to the actual outcome of the investment chosen by your leader.

You will be paid up to 10 ECU for your answer to the question randomly chosen by the computer. Your payment will be based on the following formula:

\[
\text{Your payment} = 10 \times \left[ 1 - \left( \frac{x - \text{your guess}}{100} \right)^2 \right]
\]

where \( x \) takes the value of 100 if your leader chose Investment X, and \( x \) takes the value of 0 if your leader chose Investment Y.

The formula is designed such that you can expect to receive the highest payoff if you report your true belief about the investment decision of your leader. As you can see, the closer your guess is to the actual decision made by your leader, the higher is your payoff.
Experiment 2 (Stage 3) – Treatment BSR

After you have made your investment decision, you will be asked to predict which investment your leader has chosen, assuming that someone else in your group has been assigned the role of the leader.

Specifically, we would like to know how likely it is in your opinion that the leader has chosen Investment X. *Suppose there were 100 people in the position the leader is in now. How many of them do you think would choose Investment X?*

You will need to choose a number between 0 and 100. A **higher** number means that you think the leader is **more** likely to have chosen Investment X.

The specific questions you will be asked are listed below.

**Question 1**

Suppose there were 100 people in the position the leader is in now. How many of them do you think would choose Investment X?

In Question 2, you are given additional information. You are asked to evaluate the same question with this additional information. Specifically, you should consider whether your guess of the leader’s decision will be different, given that you know the outcome of the investment chosen by your leader.

**Question 2**

*Suppose you are informed that the investment chosen by your leader has succeeded, and you have therefore received the high payoff.*

Now consider whether your guess will be higher than, lower than, or the same as the one you stated in Question 1. That is, suppose there were 100 people in the position the leader is in now. Given an outcome of high payoff, how many of them do you think have chosen Investment X?

*Suppose you are informed that the investment chosen by your leader has failed, and you have therefore received the low payoff.*

Now consider whether your guess will be higher than, lower than, or the same as the one you stated in Question 1. That is, suppose there were 100 people in the position the leader is in now. Given an outcome of low payoff, how many of them do you think have chosen Investment X?
The computer will randomly select one of these two questions and you will be paid for your response to this question. If Question 2 is chosen for payment, then you will be paid for your answer to the scenario that corresponds to the actual outcome of the investment chosen by your leader.

The section below describes how your payoff in Stage 3 will be determined. This procedure has been used in many other studies. We explain the procedure in detail, but what is most important is that this payoff structure is designed such that it is in your best interest to report your true belief about your leader’s decision.

Your payment for the question randomly chosen by the computer is determined as follows. You will receive 10 ECU with some chance. Your chance of receiving 10 ECU depends on your answer and the leader’s decision. The closer your guess is to the actual decision made by your leader, the higher is your chance of receiving the fixed payment of 10 ECU.

Specifically, your chance of receiving 10 ECU is determined by the following formula:

\[
\text{Chance of receiving 10 ECU} = \left[ 1 - \left( \frac{x - \text{your guess}}{100} \right)^2 \right] \times 100.
\]

\(x\) takes the value of 100 if your leader chose Investment X, and \(x\) takes the value of 0 if your leader chose Investment Y.

To illustrate, suppose your leader has chosen Investment X. This means that \(x = 100\) in the formula above, and your chance of receiving 10 ECU will be higher if your guess is higher. If you state 100 as your guess that the leader has chosen Investment X, then your chance of receiving 10 ECU will be \([1 - \left( \frac{100-100}{100} \right)^2] \times 100 = 100\). On the other hand, suppose your leader has chosen Investment Y instead, while your guess remains at 100. This means that \(x = 0\) in the formula above, and your chance of receiving 10 ECU will be \([1 - \left( \frac{0-100}{100} \right)^2] \times 100 = 0\).

Here is another example:

**Example 4.** Suppose you guess 70 as the chance that your leader has chosen Investment X for the group. At the end of the experiment, the computer reveals that your leader has chosen Investment X for the group. Hence, your chance of receiving 10 ECU will be \([1 - \left( \frac{100-70}{100} \right)^2] \times 100 = 91\).
To determine whether you receive 10 ECU, the computer will randomly draw a number between 0 and 100 (including decimal points). If the number drawn by the computer is less than or equal to your chance of receiving 10 ECU as determined by the formula above, then you will receive 10 ECU. Otherwise, you will receive 0 ECU. Hence, in Example 4 above, if the number randomly drawn by the computer is less than or equal to 91, then you will receive 10 ECU. Otherwise, you will receive 0 ECU.
B.2 Additional analyses of interim beliefs

In this section, we examine the impact of the belief elicitation mechanism on the reported interim beliefs in greater detail.

Table B.1 presents OLS estimates for the regressions of interim beliefs against the main treatment variable (i.e., the use of the QSR vs. BSR) and the appointment mechanisms in the leadership task. In all specifications, treatment BSR and appointment mechanism RA are the reference treatments. In the regression analyses, we also control for the subjects’ behavior in the risk game, Game 1, and order effects.

Both columns (1) and (2) of Table B.1 reveal that the reported interim beliefs are not statistically significantly different between treatments BSR and QSR (p-values = 0.277 and 0.296 in columns 1 and 2, respectively). Moreover, across both treatments, the leadership appointment mechanisms have a statistically significant impact on the reported interim beliefs.

When we interact the belief elicitation mechanism with the leadership appointment mechanism (column 3), none of the interaction terms are statistically significant (interaction terms of treatment QSR with: (i) LA: p-value = 0.443; (ii) HA: p-value = 0.571; and (iii) GA: p-value = 0.400). This suggests that there is no statistically significant difference in the impact of the leadership appointment mechanism on the reported interim beliefs between treatments BSR and QSR.
Table B.1: Regression of members’ interim beliefs (pooled)

<table>
<thead>
<tr>
<th>Variables</th>
<th>Dependent variable: Interim belief</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td>Treatment QSR</td>
<td>-1.941</td>
</tr>
<tr>
<td></td>
<td>(1.783)</td>
</tr>
<tr>
<td>Appointment mechanism LA</td>
<td>-14.157***</td>
</tr>
<tr>
<td></td>
<td>(1.017)</td>
</tr>
<tr>
<td>Appointment mechanism HA</td>
<td>9.763***</td>
</tr>
<tr>
<td></td>
<td>(0.949)</td>
</tr>
<tr>
<td>Appointment mechanism GA</td>
<td>3.363***</td>
</tr>
<tr>
<td></td>
<td>(1.007)</td>
</tr>
<tr>
<td>Game 1</td>
<td>-3.143***</td>
</tr>
<tr>
<td></td>
<td>(0.749)</td>
</tr>
<tr>
<td>% endowment in RG</td>
<td>-0.078**</td>
</tr>
<tr>
<td></td>
<td>(0.032)</td>
</tr>
<tr>
<td>Chooses high effort as leader</td>
<td>25.368***</td>
</tr>
<tr>
<td></td>
<td>(1.388)</td>
</tr>
<tr>
<td>Treatment QSR × mechanism LA</td>
<td>-1.499</td>
</tr>
<tr>
<td></td>
<td>(1.953)</td>
</tr>
<tr>
<td>Treatment QSR × mechanism HA</td>
<td>-1.024</td>
</tr>
<tr>
<td></td>
<td>(1.804)</td>
</tr>
<tr>
<td>Treatment QSR × mechanism GA</td>
<td>1.570</td>
</tr>
<tr>
<td></td>
<td>(1.864)</td>
</tr>
<tr>
<td>Constant</td>
<td>53.415***</td>
</tr>
<tr>
<td></td>
<td>(3.012)</td>
</tr>
<tr>
<td>Observations</td>
<td>3,060</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.135</td>
</tr>
<tr>
<td>Order effects</td>
<td>Y</td>
</tr>
</tbody>
</table>

Robust standard errors clustered at the subject level in parentheses. For all regressions, treatment BSR is the reference treatment, and appointment mechanism RA is the reference mechanism.

*** p<0.01, ** p<0.05, * p<0.10.
Appendix C

Appendices for Chapter 4

C.1 Experimental instructions

This section contains the instructions. Task 1 is the risk task which was common to all treatments. Task 2 is the contest task which differed between treatments. We provide the instructions for Task 2 which were used in treatments B, ED, P, and A1A2. Instructions for the other treatments were appropriately adjusted and are available from the authors upon request.
Instructions

Thank you for agreeing to take part in this study which is funded by the Australian Research Council. Please read the following instructions carefully. A clear understanding of the instructions will help you make better decisions and increase your earnings from the experiment.

The experiment consists of two separate tasks: Task 1 and Task 2. You will receive detailed instructions about each of the tasks before you participate in them. Note that the tasks are independent of each other. Your earnings will depend on the decisions you make in one of the two tasks. In other words, you will be paid for the decisions you make in either Task 1 or Task 2. This implies that you should carefully consider all of the decisions you make in both tasks in the experiment as they may determine your earnings. Whether you will be paid for Task 1 or Task 2 will be randomly determined at the end of the experiment using a coin toss. Your final payment in today’s experiment will also include a $10 participation fee.

During the experiment, we will be using Experimental Currency Units (ECU). At the end of the experiment, we will convert the amount you earn into Australian Dollars (AUD) using the following conversion rate: 10 ECU = 1 AUD.

All of the decisions you make in this experiment will remain anonymous. At the end of Task 2, you will be asked to fill out a brief questionnaire asking you some general questions.

Please do not talk to one another during the experiment. If you have any questions, please raise your hand and we will come over to answer your questions privately.
Task 1

In this task, you will be given an endowment of 50 ECU. You have the opportunity to invest any portion of this amount (i.e., any amount between 0 and 50).

The investment can either fail or succeed. There is an equal chance that the investment will fail or succeed. If the investment fails, you will lose the amount you invested. If the investment succeeds, you will receive three times the amount invested. Whether an investment fails or succeeds will be determined by the computer.

You will be given an on-screen calculator that you can use to calculate the amount you would be paid if your investment is successful and if it is not. Figure C.1 shows the decision screen for this task, where the calculator is on the left panel. You can submit your actual investment decision at the bottom of the screen. After you choose how much you wish to invest, please press the submit button.

![Figure C.1: Decision Screen for Task 1](image)

If this task is picked for payment, the computer will determine randomly whether your investment succeeds or fails.
Here are some examples:

1. Suppose you choose to invest nothing. You will get 50 ECU for sure from this task if this task is chosen for payment.
2. Suppose you choose to invest all of the endowment, i.e., 50 ECU. If the investment succeeds, you get 150 ECU. If the investment fails, you get 0.
3. Suppose you choose to invest 40 ECU. If the investment succeeds, you get 130 ECU ($40 \times 3 + 10 = 130$). If it fails, you get 10 ECU.

If you have any questions, please raise your hand. If not, please proceed to make your decision on the computer screen.
Task 2 (Treatment B)

In this task, the computer will randomly match you with one other person in the room. Hence, you will participate in Task 2 in groups of two. No one will ever be informed of the identity of the other individual in their group.

Each group member will have one of two possible roles: Player A or Player B. At the beginning of the task, you will be informed whether you have been assigned the role of Player A or Player B. Each individual in your group has an equal chance of being assigned the role of Player A or Player B.

Player A and Player B will be given the opportunity to participate in an activity. The activity is the same for both players. Player A and Player B will each be presented with a series of words. They will be asked to encode these words by substituting the letters of the alphabet with numbers using an encoding table which will be provided on their screens. Every time a word is encoded correctly, the encoding table will change.

![Figure C.2: Encoding Activity for Task 2](image)

Once you encode a word correctly, the computer will prompt you with another word which you will be asked to encode. Once you encode that word, you will be given another word and so on. This process will continue for 10 minutes (600 seconds). Player A and Player B will be given the same words to encode in the same sequence.
Figure C.2 shows an example of a word (ELPPA) to be encoded. The table given on the screen shows that E=6, L=11, P=15, P=15, and A=2.

There are two prizes, Prize 1 and Prize 2. These two prizes will be allocated to Player A and Player B according to the total number of words correctly encoded by the players.

The number of words encoded correctly by the players in the 10-minute period will determine who receives Prize 1. The player who does not receive Prize 1 will receive Prize 2. **Prize 1 is worth 150 ECU. Prize 2 is worth 150 ECU.**

Given the number of words correctly encoded by Player A and Player B, Prize 1 will be allocated in the following manner.

You can never guarantee yourself Prize 1. However, by encoding more words correctly, you can increase your chance of receiving Prize 1. The more words you encode relative to the other player, the greater your chance of receiving Prize 1. The more words the other player encodes, the less likely you are to receive Prize 1. Specifically, your chance of receiving Prize 1 is given by the following expression.

\[
\text{Your chance of receiving Prize 1} = \frac{\# \text{ words you encode correctly}}{\# \text{ words you encode correctly} + \# \text{ words the other player encodes correctly}}
\]

You can think of it in the following way. For example, if you encode 10 words correctly in total and the other player encodes 20 words correctly in total, then you will receive 10 lottery tickets while the other participant will receive 20 lottery tickets. At the end of Task 2, the computer randomly draws one of the 30 lottery tickets received by you and the other player. The owner of the drawn ticket receives Prize 1. In this example, your chance of receiving Prize 1 is 10 / (10 + 20) = 0.33 and the other participant’s chance of receiving Prize 1 is 20 / (10 + 20) = 0.67. Hence, your chance of receiving Prize 1 depends on the number of lottery tickets received by you and the other participant. The player who does not receive Prize 1 will receive Prize 2. If both Player A and Player B encode the same number of words correctly, then the computer will allocate the two prizes randomly. Each player has the same chance of receiving Prize 1.

If you have any questions, please raise your hand. Otherwise, please proceed to answer the questions on the next page. The purpose of these questions is to make sure that you understand the task. Any unclear points will be explained by the experimenter. Once you have answered all the questions, please raise your hand and one of the experimenters will come to check your answers. We will proceed with the task after checking everyone’s answers.
Practice Questions for Task 2

(Please note that the numbers in the following questions are for illustration purposes only. The questions aim to help you understand the task in a better way and should not be used as a guide for decision-making in the experiment.)

1. Assume that you encode 1 word and your opponent encodes 9 words. What is your chance of receiving Prize 1?
   (a) 1 / 9
   (b) 1 / 10
   (c) 1 / 8
   (d) 8 / 9
   (e) 9 / 10

2. Assume that you encode 9 words and your opponent encodes 1 word. What is your chance of receiving Prize 1?
   (a) 1 / 9
   (b) 1 / 10
   (c) 8 / 9
   (d) 9 / 10
   (e) 9 / 9

3. Assume that Player A encodes 45 words. State whether each of the following statements is true or false.
   (i) If Player B encodes 0 words, then Player B will definitely receive Prize 2 which is worth 150 ECU.
   (ii) If Player B encodes 1 word, then Player B will definitely receive Prize 2 which is worth 150 ECU.
   (iii) If Player B encodes 45 words, then each player will receive Prize 2 which is worth 150 ECU.
Important summary information:

- Task 2 is played in groups of two. The matchings are anonymous. You will not be told which of the participants in this room are assigned to which group.
- Player A and Player B will be given the opportunity to participate in an activity to encode words using the codes given on the screen.
- There are two prizes in this task – Prize 1 and Prize 2. You can never guarantee yourself Prize 1. However, by encoding more words correctly, you can increase your chance of receiving Prize 1. The player who does not receive Prize 1 receives Prize 2. Prize 1 is worth 150 ECU. Prize 2 is worth 150 ECU.

If you have no further questions, we will now begin a practice round of the activity. The practice round will last for 3 minutes (180 seconds). The purpose of the practice round is for you to familiarise yourself with the activity. The number of words you encode during this practice round will not affect your payment in today’s experiment.

Once you have completed the practice round, the computer will reveal the information about your assigned role for Task 2. Task 2 will then begin.
Task 2 (Treatment ED)

In this task, the computer will randomly match you with one other person in the room. Hence, you will participate in Task 2 in groups of two. No one will ever be informed of the identity of the other individual in their group.

Each group member will have one of two possible roles: Player A or Player B. At the beginning of the task, you will be informed whether you have been assigned the role of Player A or Player B. Each individual in your group has an equal chance of being assigned the role of Player A or Player B.

Player A and Player B will be given the opportunity to participate in an activity. The activity is the same for both players. Player A and Player B will each be presented with a series of words. They will be asked to encode these words by substituting the letters of the alphabet with numbers using an encoding table which will be provided on their screens. Every time a word is encoded correctly, the encoding table will change.

Once you encode a word correctly, the computer will prompt you with another word which you will be asked to encode. Once you encode that word, you will be given another word and so on. This process will continue for 10 minutes (600 seconds). Player A and Player B will be given the same words to encode in the same sequence.
Figure C.3 shows an example of a word (ELPPA) to be encoded. The table given on the screen shows that E=6, L=11, P=15, P=15, and A=2.

There are two prizes, Prize 1 and Prize 2. These two prizes will be allocated to Player A and Player B according to the total number of words correctly encoded by the players.

The number of words encoded correctly by the players in the 10-minute period will determine who receives Prize 1. The player who does not receive Prize 1 will receive Prize 2. **Prize 1 is worth 150 ECU. Prize 2 is worth 150 ECU.**

Given the number of words correctly encoded by Player A and Player B, Prize 1 will be allocated in the following manner.

You can never guarantee yourself Prize 1. However, by encoding more words correctly, you can increase your chance of receiving Prize 1. The more words you encode relative to the other player, the greater your chance of receiving Prize 1. The more words the other player encodes, the less likely you are to receive Prize 1. Specifically, your chance of receiving Prize 1 is given by the following expression.

\[
\text{Your chance of receiving Prize 1} = \frac{\# \text{ words you encode correctly}}{\# \text{ words you encode correctly} + \# \text{ words the other player encodes correctly}}
\]

You can think of it in the following way. For example, if you encode 10 words correctly in total and the other player encodes 20 words correctly in total, then you will receive 10 lottery tickets while the other participant will receive 20 lottery tickets. At the end of Task 2, the computer randomly draws one of the 30 lottery tickets received by you and the other player. The owner of the drawn ticket receives Prize 1. In this example, your chance of receiving Prize 1 is 10 / (10 + 20) = 0.33 and the other participant’s chance of receiving Prize 1 is 20 / (10 + 20) = 0.67. Hence, your chance of receiving Prize 1 depends on the number of lottery tickets received by you and the other participant. The player who does not receive Prize 1 will receive Prize 2. If both Player A and Player B encode the same number of words correctly, then the computer will allocate the two prizes randomly. Each player has the same chance of receiving Prize 1.
The task is scheduled for 10 minutes. However, if you wish to stop the task early and leave the experiment, you can. At any time during the experiment, if you wish to end the task before the 10 minutes are over, you can press the “Finish Experiment” button at the top right-hand corner of the screen displayed in Figure C.3. Once you press the button, the encoding task is blocked and you cannot continue working on the task. Figure C.4 shows the screen which will be displayed when you press the “Finish Experiment” button. You will be asked to confirm whether you want to leave the experiment early or to resume the task.

![Confirmation Screen for Task 2](image)

Figure C.4: Confirmation Screen for Task 2

If you choose to leave the experiment early, you will be shown your payment summary for today’s experiment. Remember that the computer will randomly determine which of Task 1 or Task 2 you will be paid for. If Task 2 is chosen for payment, you will receive 150 ECU for this task since Prize 1 and Prize 2 are equal. However, you will not be told whether you have received Prize 1 or Prize 2. Once you have finished reviewing your payment summary, please raise your hand and wait for an experimenter to activate the questionnaire for you to complete. After you have completed the questionnaire, please raise your hand and wait for the experimenter to call you to receive your payment.

If you have any questions, please raise your hand. Otherwise, please proceed to answer the questions on the next page. The purpose of these questions is to make sure that you understand the task. Any unclear points will be explained by the experimenter. Once you have answered all the questions, please raise your hand and one of the experimenters will come to check your answers. We will proceed with the task after checking everyone’s answers.
Practice Questions for Task 2

(Please note that the numbers in the following questions are for illustration purposes only. The questions aim to help you understand the task in a better way and should not be used as a guide for decision-making in the experiment.)

1. Assume that you encode 1 word and your opponent encodes 9 words. What is your chance of receiving Prize 1?
   (a) 1 / 9
   (b) 1 / 10
   (c) 1 / 8
   (d) 8 / 9
   (e) 9 / 10

2. Assume that you encode 9 words and your opponent encodes 1 word. What is your chance of receiving Prize 1?
   (a) 1 / 9
   (b) 1 / 10
   (c) 8 / 9
   (d) 9 / 10
   (e) 9 / 9

3. Assume that Player A encodes 45 words. State whether each of the following statements is true or false.
   (i) If Player B encodes 0 words, then Player B will definitely receive Prize 2 which is worth 150 ECU.
   (ii) If Player B encodes 1 word, then Player B will definitely receive Prize 2 which is worth 150 ECU.
   (iii) If Player B encodes 45 words, then each player will receive Prize 2 which is worth 150 ECU.

4. Assume that you choose to end the task after 3 minutes. What would be the total payment you receive for this task?
   (a) 0 ECU
   (b) 45 ECU
   (c) 75 ECU
   (d) 105 ECU
   (e) 150 ECU
Important summary information:

- Task 2 is played in groups of two. The matchings are anonymous. You will not be told which of the participants in this room are assigned to which group.
- Player A and Player B will be given the opportunity to participate in an activity to encode words using the codes given on the screen.
- There are two prizes in this task – Prize 1 and Prize 2. You can never guarantee yourself Prize 1. However, by encoding more words correctly, you can increase your chance of receiving Prize 1. The player who does not receive Prize 1 receives Prize 2. Prize 1 is worth 150 ECU. Prize 2 is worth 150 ECU.
- In addition, you have the opportunity to end the experiment and leave at any time during the task. If Task 2 is chosen for payment, you will receive 150 ECU for this task even if you choose to end the task before the 10 minutes are over. However, you will not be told whether you have received Prize 1 or Prize 2.

If you have no further questions, we will now begin a practice round of the activity. The practice round will last for 3 minutes (180 seconds). The purpose of the practice round is for you to familiarise yourself with the activity. The number of words you encode during this practice round will not affect your payment in today’s experiment. You are also not allowed to leave the experiment early during the practice round.

Once you have completed the practice round, the computer will reveal the information about your assigned role for Task 2. Task 2 will then begin.
Task 2 (Treatment P)

In this task, the computer will randomly match you with one other person in the room. Hence, you will participate in Task 2 in groups of two. No one will ever be informed of the identity of the other individual in their group.

Each group member will have one of two possible roles: Player A or Player B. At the beginning of the task, you will be informed whether you have been assigned the role of Player A or Player B. Each individual in your group has an equal chance of being assigned the role of Player A or Player B.

Player A and Player B will be given the opportunity to participate in an activity. The activity is the same for both players. Player A and Player B will each be presented with a series of words. They will be asked to encode these words by substituting the letters of the alphabet with numbers using an encoding table which will be provided on their screens. Every time a word is encoded correctly, the encoding table will change.

Once you encode a word correctly, the computer will prompt you with another word which you will be asked to encode. Once you encode that word, you will be given another word and so on. This process will continue for 10 minutes (600 seconds). Player A and Player B will be given the same words to encode in the same sequence.
Figure C.5 shows an example of a word (ELPPA) to be encoded. The table given on the screen shows that $E=6$, $L=11$, $P=15$, $P=15$, and $A=2$.

There are two prizes, Prize 1 and Prize 2. These two prizes will be allocated to Player A and Player B according to the total number of words correctly encoded by the players.

The number of words encoded correctly by the players in the 10-minute period will determine who receives Prize 1. The player who does not receive Prize 1 will receive Prize 2. **Prize 1 is worth 150 ECU. Prize 2 is worth 150 ECU.**

Given the number of words correctly encoded by Player A and Player B, Prize 1 will be allocated in the following manner.

You can never guarantee yourself Prize 1. However, by encoding more words correctly, you can increase your chance of receiving Prize 1. The more words you encode relative to the other player, the greater your chance of receiving Prize 1. The more words the other player encodes, the less likely you are to receive Prize 1. Specifically, your chance of receiving Prize 1 is given by the following expression.

$$\text{Your chance of receiving Prize 1} = \frac{\text{# words you encode correctly}}{\text{# words you encode correctly} + \text{# words the other player encodes correctly}}$$

You can think of it in the following way. For example, if you encode 10 words correctly in total and the other player encodes 20 words correctly in total, then you will receive 10 lottery tickets while the other participant will receive 20 lottery tickets. At the end of Task 2, the computer randomly draws one of the 30 lottery tickets received by you and the other player. The owner of the drawn ticket receives Prize 1. In this example, your chance of receiving Prize 1 is $10 / (10 + 20) = 0.33$ and the other participant’s chance of receiving Prize 1 is $20 / (10 + 20) = 0.67$. Hence, your chance of receiving Prize 1 depends on the number of lottery tickets received by you and the other participant. The player who does not receive Prize 1 will receive Prize 2. If both Player A and Player B encode the same number of words correctly, then the computer will allocate the two prizes randomly. Each player has the same chance of receiving Prize 1.
At any time during the experiment, if you wish to take a break, you can press the “time-out” button at the top right-hand corner of the screen displayed in Figure C.5. During the time-out, the encoding task is blocked and you cannot continue working on the task. Figure C.6 shows the screen which will be displayed when you take a time-out. If you wish to continue with the task, you can click the “resume” button in the middle of the screen. You can take as many time-outs as you want during the task.

![Time-Out Screen for Task 2](image)

Figure C.6: Time-Out Screen for Task 2

In addition to receiving either Prize 1 or Prize 2, for every minute you spend in time-out, you will earn 5 ECU. Hence, if you spend all of the 10 minutes on the “time-out” screen, you will receive 50 ECU. This means that for every second you spend on the encoding task, a proportional amount will be deducted from the total possible sum of 50 ECU. At any point during the task, you can see the amount you would receive if you click the time-out button and stay out for the remainder of the task displayed at the top right-hand corner of the screen (as shown in Figures C.5 and C.6).

If you have any questions, please raise your hand. Otherwise, please proceed to answer the questions on the next page. The purpose of these questions is to make sure that you understand the task. Any unclear points will be explained by the experimenter. Once you have answered all the questions, please raise your hand and one of the experimenters will come to check your answers. We will proceed with the task after checking everyone’s answers.
Practice Questions for Task 2

(Please note that the numbers in the following questions are for illustration purposes only. The questions aim to help you understand the task in a better way and should not be used as a guide for decision-making in the experiment.)

1. Assume that you encode 1 word and your opponent encodes 9 words. What is your chance of receiving Prize 1?
   (a) 1 / 9  
   (b) 1 / 10  
   (c) 1 / 8  
   (d) 8 / 9  
   (e) 9 / 10

2. Assume that you encode 9 words and your opponent encodes 1 word. What is your chance of receiving Prize 1?
   (a) 1 / 9  
   (b) 1 / 10  
   (c) 8 / 9  
   (d) 9 / 10  
   (e) 9 / 9

3. Assume that Player A encodes 45 words. State whether each of the following statements is true or false.
   (i) If Player B encodes 0 words, then Player B will definitely receive Prize 2 which is worth 150 ECU.  
   (ii) If Player B encodes 1 word, then Player B will definitely receive Prize 2 which is worth 150 ECU.  
   (iii) If Player B encodes 45 words, then each player will receive Prize 2 which is worth 150 ECU.

4. Assume that you spend 3 minutes on the encoding task and the remaining time on the “time-out” screen. What would be the total payment you receive for this task?
   (a) 150 ECU  
   (b) 153 ECU  
   (c) 165 ECU  
   (d) 185 ECU  
   (e) 200 ECU
Important summary information:

- Task 2 is played in groups of two. The matchings are anonymous. You will not be told which of the participants in this room are assigned to which group.
- Player A and Player B will be given the opportunity to participate in an activity to encode words using the codes given on the screen.
- There are two prizes in this task – Prize 1 and Prize 2. You can never guarantee yourself Prize 1. However, by encoding more words correctly, you can increase your chance of receiving Prize 1. The player who does not receive Prize 1 receives Prize 2. Prize 1 is worth 150 ECU. Prize 2 is worth 150 ECU.
- In addition, your payment depends on how much time you spend on the “time-out” screen. For every minute you spend in time-out, you will earn 5 ECU. Hence, if you spend all of the 10 minutes on the “time-out” screen, you will receive 50 ECU.

If you have no further questions, we will now begin a practice round of the activity. The practice round will last for 3 minutes (180 seconds). The purpose of the practice round is for you to familiarise yourself with the activity. The number of words you encode during this practice round will not affect your payment in today’s experiment. There is also no time-out payment to receive from this practice round.

Once you have completed the practice round, the computer will reveal the information about your assigned role for Task 2. Task 2 will then begin.
Task 2 (Treatment A1A2)

In this task, each player will have a choice between two activities: Activity 1 or Activity 2. You may choose to do only one, both, or none of the activities. At any time during the task, you can switch between the two activities as many times as you want. Your payment from this task will be equal to the total earnings that you receive from both Activity 1 and Activity 2. You will be given 10 minutes (600 seconds) to participate in both activities.

A screenshot of both activities is shown in Figure C.7. Activity 1 is located in the top panel with a blue frame, while Activity 2 is located in the bottom panel with a red frame.

Figure C.7: Encoding Activities for Task 2
Activity 1

To participate in Activity 1, the computer will randomly match you with one other person in the room. No one will ever be informed of the identity of the other individual in their group. Each group member will have one of two possible roles: Player A or Player B. At the beginning of the task, you will be informed whether you have been assigned the role of Player A or Player B. Each individual in your group has an equal chance of being assigned the role of Player A or Player B.

Player A and Player B will each be presented with a series of words in the same order. They will be asked to encode these words by substituting the letters of the alphabet with numbers using an encoding table which will be provided on their screens. Every time a word is encoded correctly, the encoding table will change.

As an example, Figure C.7 shows that in Activity 1, the word to be encoded is ELPPA, and the table given on the screen shows that E=6, L=11, P=15, P=15, and A=2.

Once you encode a word correctly, the computer will prompt you with another word which you will be asked to encode. Once you encode that word, you will be given another word and so on.

There are two prizes, Prize 1 and Prize 2. These two prizes will be allocated to Player A and Player B according to the total number of words correctly encoded by the players.

The number of words encoded correctly by the players in the 10-minute period will determine who receives Prize 1. The player who does not receive Prize 1 will receive Prize 2 in Activity 1. Prize 1 is worth 150 ECU. Prize 2 is worth 150 ECU.

Given the number of words correctly encoded by Player A and Player B in Activity 1, Prize 1 will be allocated in the following manner.
You can never guarantee yourself Prize 1. However, by encoding more words correctly in Activity 1, you can increase your chance of receiving Prize 1. The more words you encode relative to the other player, the greater your chance of receiving Prize 1. The more words the other player encodes, the less likely you are to receive Prize 1. Specifically, your chance of receiving Prize 1 is given by the following expression.

\[
\text{Your chance of receiving Prize 1} = \frac{\text{# words you encode in Activity 1}}{\text{# words you encode in Activity 1} + \text{# words the other player encodes in Activity 1}}
\]

You can think of it in the following way. For example, if you encode 10 words correctly in total and the other player encodes 20 words correctly in total, then you will receive 10 lottery tickets while the other participant will receive 20 lottery tickets. At the end of Task 2, the computer randomly draws one of the 30 lottery tickets received by you and the other player. The owner of the drawn ticket receives Prize 1. In this example, your chance of receiving Prize 1 is \( \frac{10}{10 + 20} = 0.33 \) and the other participant’s chance of receiving Prize 1 is \( \frac{20}{10 + 20} = 0.67 \). Hence, your chance of receiving Prize 1 depends on the number of lottery tickets received by you and the other participant. The player who does not receive Prize 1 will receive Prize 2. If both Player A and Player B encode the same number of words correctly, then the computer will allocate the two prizes randomly. Each player has the same chance of receiving Prize 1.

Activity 2

The task in Activity 2 is the same. However, you participate in this activity individually and your payment is determined differently.

Specifically, in Activity 2, you will receive 1.25 ECU for each word that you encode correctly. Hence, if you encode 4 words correctly in Activity 2, then you will receive 5 ECU from this activity.

In both Activity 1 and Activity 2, all participants will be given the same encoding tables and the same words in the same order. However, the encoding tables and the sequence of words will be different between the two activities. As an example, Figure C.7 shows that while the word is ELPPA in Activity 1, it is REPAP in Activity 2. The table given on the screen shows that R=17, E=22, P=15, A=26, and P=15.
At the end of the experiment, if Task 2 is picked for payment, then you will receive:

Payment for Task 2

\[ = (\text{Prize 1 or Prize 2}) + (\text{number of words you encode in Activity 2} \times 1.25 \text{ ECU}). \]

If you have any questions, please raise your hand. Otherwise, please proceed to answer the questions on the next page. The purpose of these questions is to make sure that you understand the task. Any unclear points will be explained by the experimenter. Once you have answered all the questions, please raise your hand and one of the experimenters will come to check your answers. We will proceed with the task after checking everyone’s answers.
Practice Questions for Task 2

(Please note that the numbers in the following questions are for illustration purposes only. The questions aim to help you understand the task in a better way and should not be used as a guide for decision-making in the experiment.)

1. In Activity 1, assume that you encode 1 word and your opponent encodes 9 words. What is your chance of receiving Prize 1?
   (a) 1 / 9
   (b) 1 / 10
   (c) 1 / 8
   (d) 8 / 9
   (e) 9 / 10

2. In Activity 1, assume that you encode 9 words and your opponent encodes 1 word. What is your chance of receiving Prize 1?
   (a) 1 / 9
   (b) 1 / 10
   (c) 8 / 9
   (d) 9 / 10
   (e) 9 / 9

3. In Activity 1, assume that Player A encodes 45 words. State whether each of the following statements is true or false.
   (i) If Player B encodes 0 words in Activity 1, then Player B will definitely receive Prize 2 which is worth 150 ECU.
   (ii) If Player B encodes 1 word in Activity 1, then Player B will definitely receive Prize 2 which is worth 150 ECU.
   (iii) If Player B encodes 45 words in Activity 1, then each player will receive Prize 2 which is worth 150 ECU.
4. Suppose you encode 20 words correctly in Activity 1, and 40 words correctly in Activity 2. Suppose also that Player B encodes 20 words correctly in Activity 1. What would be the total payment you can receive from this task?
(a) 40 ECU  
(b) 60 ECU  
(c) 75 ECU  
(d) 150 ECU  
(e) 200 ECU  

5. Suppose you encode 0 words correctly in Activity 1, and 40 words correctly in Activity 2. Suppose also that Player B encodes 20 words correctly in Activity 1. What would be the total payment you can receive from this task?
(a) 40 ECU  
(b) 60 ECU  
(c) 75 ECU  
(d) 150 ECU  
(e) 200 ECU
Important summary information:

- In Task 2, there are two activities. You may choose to do only one, both, or none of the activities. At any time during the task, you can switch between the two activities as many times as you want.
- Activity 1 is played in groups of two. The matchings are anonymous. You will not be told which of the participants in this room are assigned to which group.
- Activity 2 is played individually.
- In each activity, you will be asked to encode words using the codes given on the screen.
- Your payment from this task will be equal to the total earnings that you receive from both Activity 1 and Activity 2.
- In Activity 1, there are two prizes—Prize 1 and Prize 2. You can never guarantee yourself Prize 1. However, by encoding more words correctly, you can increase your chance of receiving Prize 1. The player who does not receive Prize 1 receives Prize 2. Prize 1 is worth 150 ECU. Prize 2 is worth 150 ECU.
- In Activity 2, you will receive 1.25 ECU for each word that you encode correctly.

If you have no further questions, we will now begin a practice round of the activity. The practice round will last for 3 minutes (180 seconds). The purpose of the practice round is for you to familiarise yourself with the activity. The number of words you encode during this practice round will not affect your payment in today’s experiment.

Once you have completed the practice round, the computer will reveal the information about your assigned role for Task 2. Task 2 will then begin.
C.2 Additional analyses

In this section, we present the analysis including three additional treatments: (i) Slider (S); (ii) 20-Minute (20-Min); and (iii) Effort Choice (EC). We show that the main results in the paper continue to hold after the inclusion of the data from these treatments. Table C.1 presents a summary of all the treatments including these three additional treatments. Figure C.8 presents histograms of the distribution of effort in these additional treatments.

![Histograms](image)

(a) Treatment S

(b) Treatment 20-Min

(c) Treatment EC

Figure C.8: Overall effort choices by treatment (S, 20-Min, and EC)

Data from these three treatments are pooled together with the data from Experiment 2, and Table C.2 presents the participation rate and average effort per minute conditional on participation in these treatments. Table C.3 presents the results of OLS and quantile (median) regressions of participation (column 1) and effort per minute (columns 2 and 3) for Experiment 2 including these additional treatments.
A comparison of treatments S and 20-Min to treatment B reveals that, in the absence of monetary incentives, both participation and effort per minute remain very high even when a different task is used or when the duration of the task is increased. 94% and 92% of the subjects exert positive effort in treatments S and 20-Min, respectively. These participation rates are not significantly different from that in treatment B (Fisher’s exact tests, B vs. S: p-value = 0.678; B vs. 20-Min: p-value = 0.433). The average effort per minute in treatment 20-Min is 4.22, which is also not significantly different to that in treatment B (Wilcoxon rank-sum test, B vs. 20-Min: p-value = 0.959). The regression estimates in Table C.3 are also consistent with these findings.

Finally, we find that the behavior of the subjects in treatment EC is largely similar to that in Activity 1 of treatment A1A2. 36% of subjects exert positive effort in treatment EC, and this is not statistically different to that in treatment A1A2 (Fisher’s exact test, A1A2 vs. EC: p-value = 1.000). This suggests that competitiveness is a key driver of effort for some subjects who choose to exert effort even in the absence of monetary prizes. This result is also consistent with Sheremeta (2010).

\footnote{We do not compare the average effort per minute between treatments B and S since the tasks used in these treatments are different.}
Table C.1: Summary of experimental treatments (including treatments S, 20-Min, and EC)

<table>
<thead>
<tr>
<th>Treatment</th>
<th># subjects</th>
<th>Prize 1 (ECU)</th>
<th>Prize 2 (ECU)</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Experiment 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline (B)</td>
<td>54</td>
<td>150</td>
<td>150</td>
<td>Baseline treatment with no incentives</td>
</tr>
<tr>
<td>Baseline – Low (BL)</td>
<td>50</td>
<td>250</td>
<td>150</td>
<td>Baseline treatment with low incentives</td>
</tr>
<tr>
<td>Baseline – High (BH)</td>
<td>52</td>
<td>600</td>
<td>150</td>
<td>Baseline treatment with high incentives</td>
</tr>
<tr>
<td>(b) Experiment 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline (B)</td>
<td>54</td>
<td>150</td>
<td>150</td>
<td>Baseline treatment</td>
</tr>
<tr>
<td>Early Departure (ED)</td>
<td>50</td>
<td>150</td>
<td>150</td>
<td>Outside option: Can leave experiment early</td>
</tr>
<tr>
<td>Pause (P)</td>
<td>56</td>
<td>150</td>
<td>150</td>
<td>Outside option: Timeout payment</td>
</tr>
<tr>
<td>Activity 1/2 (A1A2)</td>
<td>56</td>
<td>150</td>
<td>150</td>
<td>Outside option: Same activity with piece rate</td>
</tr>
<tr>
<td>Slider (S)</td>
<td>54</td>
<td>150</td>
<td>150</td>
<td>Slider task used</td>
</tr>
<tr>
<td>20-Minute (20-Min)</td>
<td>56</td>
<td>150</td>
<td>150</td>
<td>Longer duration</td>
</tr>
<tr>
<td>Effort Choice (EC)</td>
<td>56</td>
<td>150</td>
<td>150</td>
<td>No real-effort task</td>
</tr>
<tr>
<td>(c) Experiment 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Activity 1/2 (A1A2)</td>
<td>56</td>
<td>150</td>
<td>150</td>
<td>A1A2 treatment with no incentives</td>
</tr>
<tr>
<td>Activity 1/2 – Low (A1A2-L)</td>
<td>56</td>
<td>250</td>
<td>150</td>
<td>A1A2 treatment with low incentives</td>
</tr>
<tr>
<td>Activity 1/2 – High (A1A2-H)</td>
<td>52</td>
<td>600</td>
<td>150</td>
<td>A1A2 treatment with high incentives</td>
</tr>
</tbody>
</table>
Table C.2: Summary statistics (including treatments S, 20-Min, and EC)

<table>
<thead>
<tr>
<th>Treatment</th>
<th># subjects(^{(a)})</th>
<th>Participation rate</th>
<th>Average effort per minute(^{(b)})</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) <strong>Experiment 1</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline (B)</td>
<td>54</td>
<td>0.96</td>
<td>4.22</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[0.19]</td>
<td>[0.81]</td>
</tr>
<tr>
<td>Baseline – Low (BL)</td>
<td>50</td>
<td>1.00</td>
<td>4.49</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[0.00]</td>
<td>[0.56]</td>
</tr>
<tr>
<td>Baseline – High (BH)</td>
<td>52</td>
<td>1.00</td>
<td>4.33</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[0.00]</td>
<td>[0.86]</td>
</tr>
<tr>
<td>(b) <strong>Experiment 2</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline (B)</td>
<td>54</td>
<td>0.96</td>
<td>4.22</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[0.19]</td>
<td>[0.81]</td>
</tr>
<tr>
<td>Early Departure (ED)</td>
<td>50</td>
<td>0.82</td>
<td>3.27</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[0.39]</td>
<td>[1.56]</td>
</tr>
<tr>
<td>Pause (P)</td>
<td>56</td>
<td>0.54</td>
<td>0.94</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[0.50]</td>
<td>[1.12]</td>
</tr>
<tr>
<td>Activity 1/2 (A1A2)</td>
<td>53</td>
<td>0.38</td>
<td>0.54</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[0.49]</td>
<td>[0.88]</td>
</tr>
<tr>
<td>Slider (S)</td>
<td>53</td>
<td>0.94</td>
<td>7.18</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[0.23]</td>
<td>[1.92]</td>
</tr>
<tr>
<td>20-Minute (20-Min)</td>
<td>52</td>
<td>0.92</td>
<td>4.22</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[0.27]</td>
<td>[0.91]</td>
</tr>
<tr>
<td>Effort Choice (EC)</td>
<td>55</td>
<td>0.36</td>
<td>.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[0.49]</td>
<td>[.]</td>
</tr>
<tr>
<td>(c) <strong>Experiment 3</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Activity 1/2 (A1A2)</td>
<td>53</td>
<td>0.38</td>
<td>0.54</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[0.49]</td>
<td>[0.88]</td>
</tr>
<tr>
<td>Activity 1/2 – Low (A1A2-L)</td>
<td>56</td>
<td>0.98</td>
<td>2.52</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[0.13]</td>
<td>[1.74]</td>
</tr>
<tr>
<td>Activity 1/2 – High (A1A2-H)</td>
<td>52</td>
<td>0.96</td>
<td>3.27</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[0.19]</td>
<td>[1.44]</td>
</tr>
</tbody>
</table>

Sample means given. Standard deviations in brackets.
For treatments A1A2, A1A2-L, and A1A2-H, we report the participation rate and average effort per minute in Activity 1.

\(^{(a)}\) We discovered that nine subjects (three in treatment A1A2; one in treatment S; four in treatment 20-Min; and one in treatment EC) had previously participated in the same experiment. Hence, we drop these subjects in our analysis.

\(^{(b)}\) Average effort per minute is defined as the average effort by subjects who exerted a positive level of effort.
Table C.3: Regression results for Experiment 2 (including treatments S, 20-Min, and EC)

<table>
<thead>
<tr>
<th>Variables</th>
<th>Participation =1 if Effort &gt;0 (OLS)</th>
<th>Effort per minute given Effort &gt;0 (OLS)</th>
<th>Effort per minute given Effort &gt;0 (Quantile)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early Departure (ED)</td>
<td>-0.161**</td>
<td>-0.940***</td>
<td>-0.581***</td>
</tr>
<tr>
<td></td>
<td>(0.076)</td>
<td>(0.237)</td>
<td>(0.191)</td>
</tr>
<tr>
<td>Pause (P)</td>
<td>-0.431***</td>
<td>-3.255***</td>
<td>-4.002***</td>
</tr>
<tr>
<td></td>
<td>(0.073)</td>
<td>(0.252)</td>
<td>(0.203)</td>
</tr>
<tr>
<td>Slider (S)</td>
<td>-0.041</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.074)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Effort Choice (EC)</td>
<td>-0.605***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.074)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20-Minute (20-Min)</td>
<td>-0.042</td>
<td>0.032</td>
<td>0.014</td>
</tr>
<tr>
<td></td>
<td>(0.074)</td>
<td>(0.220)</td>
<td>(0.177)</td>
</tr>
<tr>
<td>Activity 1/2 (A1A2)</td>
<td>-0.590***</td>
<td>-3.720***</td>
<td>-4.156***</td>
</tr>
<tr>
<td></td>
<td>(0.074)</td>
<td>(0.296)</td>
<td>(0.238)</td>
</tr>
<tr>
<td>% invested in risk task</td>
<td>-0.174**</td>
<td>-0.278</td>
<td>0.096</td>
</tr>
<tr>
<td></td>
<td>(0.072)</td>
<td>(0.297)</td>
<td>(0.239)</td>
</tr>
<tr>
<td>Age</td>
<td>-0.002</td>
<td>0.003</td>
<td>-0.008</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.020)</td>
<td>(0.016)</td>
</tr>
<tr>
<td>Female</td>
<td>0.013</td>
<td>0.342**</td>
<td>0.121</td>
</tr>
<tr>
<td></td>
<td>(0.041)</td>
<td>(0.167)</td>
<td>(0.135)</td>
</tr>
<tr>
<td>Economics major</td>
<td>-0.021</td>
<td>-0.137</td>
<td>-0.185</td>
</tr>
<tr>
<td></td>
<td>(0.063)</td>
<td>(0.263)</td>
<td>(0.212)</td>
</tr>
<tr>
<td>Undergraduate student</td>
<td>0.087</td>
<td>0.835</td>
<td>0.773*</td>
</tr>
<tr>
<td></td>
<td>(0.136)</td>
<td>(0.559)</td>
<td>(0.450)</td>
</tr>
<tr>
<td>Graduate student</td>
<td>0.146</td>
<td>0.910*</td>
<td>0.829*</td>
</tr>
<tr>
<td></td>
<td>(0.128)</td>
<td>(0.535)</td>
<td>(0.431)</td>
</tr>
<tr>
<td>Australian</td>
<td>-0.132***</td>
<td>0.049</td>
<td>0.079</td>
</tr>
<tr>
<td></td>
<td>(0.043)</td>
<td>(0.181)</td>
<td>(0.146)</td>
</tr>
<tr>
<td># previous experiments</td>
<td>-0.001</td>
<td>0.045</td>
<td>0.017</td>
</tr>
<tr>
<td></td>
<td>(0.011)</td>
<td>(0.044)</td>
<td>(0.036)</td>
</tr>
<tr>
<td>Constant</td>
<td>1.064***</td>
<td>3.215***</td>
<td>3.556***</td>
</tr>
<tr>
<td></td>
<td>(0.209)</td>
<td>(0.818)</td>
<td>(0.659)</td>
</tr>
<tr>
<td>Observations</td>
<td>373</td>
<td>191</td>
<td>191</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.342</td>
<td>0.658</td>
<td>-</td>
</tr>
</tbody>
</table>

Ordinary least squares estimates given in (1) and (2), and quantile (median) regression estimates given in (3). Standard errors in parentheses. For all regressions, treatment B is the reference treatment.

In all regressions, the controls are treatment variables and subjects’ characteristics, which include the percentage of the endowment invested in the risk task, age, gender, study level at University, whether the subject is pursuing a major in economics, whether the subject is Australian, and previous experience with economic experiments.

For treatment A1A2, we consider participation and effort per minute in Activity 1.

*** p<0.01, ** p<0.05, * p<0.10.
Author/s:
Koh, Boon Han

Title:
Essays in leadership and contests

Date:
2017

Persistent Link:
http://hdl.handle.net/11343/210775

File Description:
Essays in leadership and contests

Terms and Conditions:
Terms and Conditions: Copyright in works deposited in Minerva Access is retained by the copyright owner. The work may not be altered without permission from the copyright owner. Readers may only download, print and save electronic copies of whole works for their own personal non-commercial use. Any use that exceeds these limits requires permission from the copyright owner. Attribution is essential when quoting or paraphrasing from these works.