8-7-2018

Essays in Behavioral Labor and Welfare Economics

Aleksandr Alekseev

ABSTRACT

ESSAYS IN BEHAVIORAL LABOR AND WELFARE ECONOMICS

BY

ALEKSANDR ALEKSEEV

AUGUST 2018

Committee Chair: Dr. James C. Cox

Major Department: Economics

In my dissertation, I study three questions in behavioral labor and welfare economics using experimental methods: what is the role of task difficulty on effort, how to measure ability and motivation in a more rigorous way, and what are the economic consequences of stochastic choice in a risk setting.

In the first chapter, I study the effect of task difficulty on workers' effort and compare it to the effect of monetary rewards in a tightly controlled laboratory experiment. I find that task difficulty has an inverse-U effect on effort, and that this effect is quantitatively large when compared to the effect of conditional rewards. Difficulty acts as an important mediator of monetary rewards: they are most effective at the medium level of difficulty. I show that the inverse-U pattern of effort response to difficulty is not consistent with the Expected Utility model but is consistent with the Rank-Dependent Utility (RDU) model that allows for probability weighting. I structurally estimate the RDU model and find that it fits the data well. These findings suggests that 1) task difficulty is a useful and costless tool to stimulate effort, 2) to elicit the maximum amount of effort, the task has to be reasonably challenging, and 3) the design of optimal incentive schemes for workers should take into account task difficulty.

In the second chapter, I develop a novel method for estimating ability and motivation from the outcomes and response time on a cognitive test. The proposed method is based on a dynamic stochastic model of optimal effort choice that features a psychologically plausible mechanism of decision-making. In a laboratory experiment, I find substantial heterogeneity among subjects in terms of their estimated ability and motivation that is partially attributed to their demographic

characteristics and preferences. Test scores turns out to be a very imprecise measure of true ability. The observed variation in test scores is mostly due to variation in motivation rather than ability. I find no association between estimated measures of ability and motivation and their self-reported counterparts. Looking at the relative importance of ability versus motivation on the success on a cognitive task, I find that motivation plays a slightly bigger role than ability.

In the third chapter, which is a joint work with Dr. Glenn W. Harrison, Dr. Morten Lau and Dr. Don Ross, we study the welfare costs of stochastic choice. Theoretical work on stochastic choice mainly focuses on the sources of choice randomness, and less on its economic consequences. We attempt to close this gap by developing a method of extracting information about the monetary costs of noise from structural estimates of preferences and choice randomness. Our method is based on allowing a degree of noise in choices in order to rationalize them by a given structural model. To illustrate the approach, we consider risky binary choices made by a sample of the general Danish population in an artefactual field experiment. The estimated welfare costs are small in terms of everyday economic activity, but they are considerable in terms of the actual stakes of the choice environment. Higher welfare costs are associated with higher age, lower education, and lower income.

ESSAYS IN BEHAVIORAL LABOR AND WELFARE ECONOMICS

BY

ALEKSANDR ALEKSEEV

A Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree

of

Doctor of Philosophy

in the

Andrew Young School of Policy Studies

of

Georgia State University

GEORGIA STATE UNIVERSITY

2018

ACCEPTANCE

This dissertation was prepared under the direction of Aleksandr G. Alekseev's Dissertation Committee. It has been approved and accepted by all members of that committee, and it has been accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Economics in the Andrew Young School of Policy Studies of Georgia State University.

Dissertation Chair:	Dr. James C. Cox
Committee:	Dr. Glenn W. Harrison
	Dr. Vjollca Sadiraj
	Dr. J. Todd Swarthout

Electronic Version Approved:

Sally Wallace, Dean
Andrew Young School of Policy Studies
Georgia State University
August, 2018

Acknowledgments

First of all, I would like to thank my doctoral advisors. I thank Jim Cox, my committee chair, for his guidance, support, and patience. I thank Glenn Harrison, who could be called my co-chair if this title existed, for his mentorship, encouragement, and constructive criticism. I spent hours in discussions with both Jim and Glenn, which helped me shape my dissertation and, more broadly, my understanding of economics and my approach to research. I thank Todd Swarthout for teaching me the practical details of conducting laboratory experiments. I thank Vjollca Sadiraj for always being supportive and encouraging.

I received helpful advice from the faculty members at GSU who were not a part of my dissertation committee. I thank Tom Mroz for his extremely insightful comments and suggestions. I thank Susan Laury for her advice on experimental design and for her guidance in the formalities of running laboratory experiments. I thank Rusty Tchernis, Andrew Feltenstein, and Ajay Subramanian for their constant support and encouragement. I thank Dan Kreisman and Garth Heutel for their help in preparing for the job market. I thank Gary Charness for being my co-author and writing me a recommendation letter. I thank Morten Lau and Don Ross for their valuable input on the third chapter of my dissertation.

My research and studying at GSU would not be as smooth and comfortable without the help of the staff. I thank Bess Blyler for being an outstanding administrator and for managing my recommendation letters. I thank Lucy Gentles, Mark Schneider, and Thomas Kelly for their professionalism and for helping me in obtaining the funding for my experiments and conference trips. I thank Kevin Ackaramongkolrotn for always being there when technical help was needed.

I appreciate the financial support provided by GSU. I thank ExCEN, and in particular Jim Cox and Lucy Gentles, for providing me with the generous support for conference travels. I thank CEAR, and in particular Glenn Harrison, Mark Schneider, and Thomas Kelly, for the generous support I received for conducting the experiment for my first chapter and for conference travels. I thank AYSPS, and in particular Cynthia Searcy, for the generous support I received to conduct the experiment for my second chapter.

I benefited a lot from participation in numerous conferences and seminars. I thank seminar participants at the European University at St. Petersburg, George Mason University, Georgia State University, and University of Chicago. I thank conference participants at the SEA meetings in Tampa, Washington DC, and New Orleans, at the ESA meetings in Richmond, San Diego, Tucson, and Dallas, and at the WEAI meeting in San Diego. I thank the organizers and participants of the Spring School in Behavioral Economics in San Diego, and in particular Uri Gneezy, Alexander Kappelen, and Bertil Tungodden. I thank John List and his research group for their hospitality during my brief visit at the University of Chicago. I thank the participants of the Experimental Methods Forum at GSU for their feedback on my work.

I am grateful to my friends at GSU and those whom I met at conferences. I thank Klajdi Bregu, Sherry Gao, and Jean Paul Rabanal for their feedback on my work, support, and engaging discussions. I thank Rhita Simorangkir for being supportive and challenging at the same time. I thank Ano Chatterjee, Chandrayee Chatterjee (the two are not related), Lily Li, Prithvi Mukherjee, Maria Sudibjo, Susan Tang, and Dustin Tracy for their help with running my experiments.

Finally, I would like to thank my family. I thank my mom, Tatiana Basko, and my grandma, Zinaida Basko, for their limitless support and for always believing in me and my success. I thank my wife, Valentina Alekseeva, for never giving up efforts to understand my work, for insightful discussions, and for help in preparations for the job market and conference presentations. I thank my aunt, Larisa Mitrofanova, my uncle, Dmitri Mitrofanov, and my cousin, Natasha Vinokurova, for always showing their support. I thank my mother-in-law, Irina Afanasyeva, my sister-in-law, Anna Gunich, and her husband, Ivan Gunich, for being supportive and showing a genuine interest in trying to understand behavioral economics.

Contents

List of Tables

List of Figures

Chapter 1

Give Me a Challenge or Give Me a Raise

1.1 Introduction

Much of the work in labor economics highlights the role of extrinsic rewards in stimulating workers' effort. Extrinsic rewards are usually effective in achieving this goal, however, their effectiveness is constrained by various psychological factors.[1] Recent work in behavioral economics suggests that there are alternative tools to stimulate effort that are, unlike extrinsic rewards, often costless to implement.[2] One such tool, task difficulty, has been advocated by researchers in psychology as the crucial element in determining workers' effort (see Gendolla, Wright, and Richter (2012) for an overview). However, its theoretical and empirical role in an economic environment remains unexplored. Understanding the role of task difficulty and its interaction with other incentive elements, such as extrinsic monetary rewards, is important for designing optimal incentive

[1] Such factors include, e.g., the crowding out of intrinsic motivation (Gneezy and Rustichini, 2000) and the choking-under-pressure (Ariely, Gneezy, Loewenstein, and Mazar, 2009; Genakos, Pagliero, and Garbi, 2015; Hickman and Metz, 2015). See Gneezy, Meier, and Rey-Biel (2011) for an excellent overview of the cases when extrinsic rewards do and do not work.

[2] For example, setting non-binding goals (Smithers, 2015; Corgnet, Gómez-Miñambres, and Hernán-González, 2015; Goerg and Kube, 2012) or revealing to workers their performance rank (Blanes i Vidal and Nossol, 2011) has been shown to boost effort at no extra cost. Loss aversion is another effective behavioral tool that simply changes the way the extrinsic rewards are provided without adding to the cost (Fryer Jr, Levitt, List, and Sadoff, 2012; Hossain and List, 2012). The role of behavioral tools in the productivity of workers is highlighted in the 2015 World Bank Report (World Bank Group, 2015).

schemes for workers.[3] This chapter seeks to accomplish two goals: to empirically study the effect of task difficulty on effort and compare its effect with the effect of monetary rewards, and to provide a theoretical explanation of the observed effect of task difficulty.

I define task difficulty as the characteristic of a task that, when increased, reduces the probability of success in the task for any given level of effort. This definition implies that to achieve a given probability of success, i.e., performance, in a difficult task requires more effort than in an easy task,[4] which corresponds well to the intuitive notion of task difficulty.[5] For example, consider a bank clerk who proofreads contracts and has to find all errors to successfully complete a task. Contracts can be either 10-page, 20-page, or 100-page in length. The clerk's task varies in difficulty: spending 1 hour on a 10-page contract is more likely to result in finding all errors than spending 1 hour on a 100-page contract, or equivalently, it requires more time to proofread a 100-page contract than a 10-page contract. How would the clerk's work effort change across the three levels of difficulty? Does the answer to this question depend on how the clerk is being paid?

Research in psychology, and in particular motivational intensity theory (Brehm and Self, 1989; Wright, 1996), suggests that effort is determined by the minimum amount of work needed to complete the task, as long as success seems possible and beneficial. This implies that effort increases with task difficulty up to a point, after which effort drops leading to an inverse-U pattern of response. In the example above, the clerk would spend more effort on a 20-page contract relative to a 10-page contract, but would probably spend little time on a 100-page contract deeming it impossible or unjustified to proofread carefully. Empirical studies in psychology provide extensive support[6] for the inverse-U relation between task difficulty and effort (Smith, Baldwin, and Christensen, 1990; Brockner, Grover, Reed, and Dewitt, 1992; Richter, Friedrich, and Gendolla,

[3]The recent Nobel prize lecture by Holmström (2017) advocates studying the whole array of incentive tools that is not limited to simple monetary rewards.

[4]Here, for simplicity, I assume a cognitive production technology with no fixed costs.

[5]Note that task difficulty is different from the cost of effort, even though the two are sometimes mixed in the literature, as in Bremzen, Khokhlova, Suvorov, and Van de Ven (2015). Cost measures disutility from exerting effort and depends both on a task and preferences of an individual, whereas difficulty is a characteristic of a task. Low difficulty does not imply low disutility: it might be easy to wash a dirty plate but few people would find it pleasant!

[6]It should be noted, however, that in many studies in psychology incentives were absent by the standards of experimental economics.

2008; Richter, 2015). Task difficulty also plays an important role in activating analytical reasoning (Alter, Oppenheimer, Epley, and Eyre, 2007) and achieving goals (Labroo and Kim, 2009).

I study the effect of task difficulty and monetary rewards on effort in a tightly controlled laboratory experiment. Subjects are presented with projects that are defined by four characteristics: unconditional reward, conditional reward, cost, and difficulty. The outcome of a project, success or failure, determines whether a subject receives a high revenue (unconditional reward plus conditional reward) or low revenue (unconditional reward only) from the project. The probability of success in a project increases with the subject's chosen effort level and decreases with the project's difficulty. Higher effort level leads to a higher monetary cost of effort that is subtracted from the project's revenue. The cost of effort also increases with the cost characteristic of a project. In choosing their effort levels subjects thus have to consider a trade-off between a higher chance of success and a lower profit (revenue minus cost). Each subject participates in a series projects that differ in their four characteristics. I then analyze the change in effort caused by the variation in these characteristics.

To establish theoretical predictions, I begin with the benchmark Expected Utility (EU) model, in which the agent chooses effort to maximize the weighted average of the outcome utilities with the weights being the probability of each outcome. I show that risk averse agents with a concave utility-of-money function and a monetary cost of effort would monotonically decrease their effort in response to higher difficulty, and that risk neutral agents would not change their effort in response to difficulty. This result is driven by the complementarity/substitutability relation between effort and money in the agent's utility function. Under the EU model, agents would increase their effort in response to higher conditional rewards and lower cost of effort, and weakly increase their effort in response to higher unconditional rewards.

Given that the EU model cannot generate the non-monotonic inverse-U pattern of effort response to difficulty, I consider an alternative Rank-Dependent Utility (RDU) model (Quiggin, 1982). In this model, the agent also chooses effort to maximize a weighted average of the outcome utilities, but unlike the EU model, the weights are outcome probabilities that are transformed

using a probability weighting function. This function, in addition to the utility function, captures the agent's risk attitude. The RDU model includes the EU model as a special case when the probability weighting function is identity. I show that allowing for the probability weighting makes it possible to generate non-monotonic responses of effort to difficulty. In particular, when the probability weighting function is inverse-S-shaped, which is a popular specification in the empirical studies (Bruhin, Fehr-Duda, and Epper, 2010; Wu and Gonzalez, 1996) despite the lack of general evidence (Wilcox, 2015b), the pattern of effort response to difficulty is U-shaped. On the other hand, in order to generate the empirically observed inverse-U-shaped pattern of response to difficulty, the probability weighting function would need to be S-shaped.

This experimental design generates a rich dataset and provides a rigorous test of the theoretical predictions. First, the use of the chosen effort framework produces unambiguously signed comparative statics effects by controlling for the supermodularity of the utility function. Second, the within-subject design controls for subject-specific characteristics when conducting the analysis of treatment effects, and allows one to study the individual heterogeneity in effort responses. Third, I control for subjects' risk preferences by using a separate risk elicitation task and also control for their observable demographic characteristics in a demographic survey.

I find that monetary incentives affect effort as predicted. Higher conditional rewards and lower cost increase effort, while unconditional rewards have only a weakly positive effect. Difficulty has an inverse-U effect on effort, which is consistent with the S-shaped probability weighting function in the RDU model. These effects hold on the aggregate level and are robust to different statistical tests, paired-samples analysis, and the analysis of *ceteris paribus* pairs. Quantitatively, the effect of increasing difficulty from the lowest to the medium level (or decreasing it from the highest to medium level) is comparable to the effect of doubling conditional rewards. Difficulty acts as an important mediator of monetary incentives. Both conditional and unconditional rewards are most effective in stimulating effort at the medium level of difficulty. Keeping difficulty at the medium level also dampens the negative effect of cost.

Subjects' effort responses to the treatment variables differ in both types and magnitudes. The distributions of the individual responses are in general well approximated by a normal distribution. The analysis of the interaction between the subjects' gender and response types shows that males are more likely to exhibit inverse-U response to difficulty than females, while females are more likely to exhibit increasing and U-shaped types of responses.

Having established the qualitative consistency of the data with the RDU model, I study its quantitative consistency by conducting a structural analysis of the model. The estimated RDU model captures the observed behavioral patterns well and confirms that the S-shaped probability weighting is needed to generate the inverse-U response to difficulty. I then re-estimate the model using data from a lottery task generated by the same subjects and find that the model estimated on this alternative task implies an *inverse*-S-shaped probability weighting function. I discuss the potential explanations of the differences in the estimates across the two tasks.

Most closely related to this design is the research by Vandegrift and Brown (2003) that studies the interactive effect of task difficulty and conditional monetary rewards on performance in a tournament environment. It finds that monetary rewards do not have any effect on performance for easy tasks, but do have a positive effect for difficult tasks.[7] The findings in the present paper confirm this result. I extend this result by studying the unconditional effect of difficulty on effort, as well as the interaction between difficulty, unconditional rewards, and cost.

A related strand of literature looks at the effect of goals on workers' effort. Corgnet, Gómez-Miñambres, and Hernán-González (2015) conduct a laboratory principal-agent experiment and find that principals tend to set challenging but attainable performance goals for agents, which increases agents' performance relative to a no-goals baseline. They also report that the effect of goal-setting on effort[8] is stronger under high monetary rewards. Smithers (2015) conducts a laboratory experiment and finds that setting exogenous goals in an addition task increases subjects'

[7] McDaniel and Rutström (2001) report that for a difficult task, increasing penalty for bad performance tends to induce more effort from subjects, though the study does not vary task difficulty.

[8] It should be noted that higher induced effort, or performance, due to higher incentives does not necessarily increase a worker's welfare. The focus on performance is motivated from the viewpoint of a principal, but not a worker.

performance, with the effect being most pronounced in the male participants. Goerg and Kube (2012) conduct a field experiment at a campus library in which subjects were paid to sort books. They report that both exogenous and endogenous goal-setting lead to higher performance. These results are similar to what the present chapter finds, however goal difficulty and task difficulty are not equivalent (Campbell and Ilgen, 1976). Apart from having a difficulty component to them, goals also have a reference point component (Heath, Larrick, and Wu, 1999). The present paper chapter the pure, i.e., without reference dependence, effect of task difficulty on effort and provides a theoretical explanation of the mechanism through which it works.

The present work ties in to the literature on the behavioral effects of monetary rewards. Hossain and List (2012) conduct a field experiment at a Chinese manufacturing firm. They randomly assign workers to one of two conditions: in one condition, workers are paid a bonus upon reaching a given performance goal; and in the other condition, workers are paid a bonus in advance and lose it if they do not reach a performance goal. Consistent with the loss aversion hypothesis, the workers' performance was higher in the second condition. Gneezy and List (2006) conduct a field experiment at a campus library in which they vary the unconditional rewards paid to their subjects for arranging books. They find that increasing the rewards has a positive but short-lived effect on the subjects' effort. A similar finding is reported by Jayaraman, Ray, and de Vèricourt (2016) who study the effort response of tea pluckers in India to an increase in their unconditional rewards caused by a change in their contracts. Hennig-Schmidt, Sadrieh, and Rockenbach (2010) show that the positive effect of unconditional rewards occurs only when workers understand the benefit of their work to the principal, which triggers positive reciprocity. The weakly positive effect of unconditional rewards reported in the present chapter is consistent with their findings.

Finally, the present work contributes to studies of the role of alternative behavioral tools in stimulating workers' effort. Charness, Cobo-Reyes, Jiménez, Lacomba, and Lagos (2012) conduct a laboratory gift-exchange experiment in which the principal can delegate the choice of wage and effort to the agent, as well as the choice of a desired effort goal. They find that delegation is profitable both for the principal and the agent, contrary to the standard economic prediction.

Blanes i Vidal and Nossol (2011) present quasi-experimental evidence from a firm in which workers compensated by piece rates receive information about their relative performance. They find that providing this information has a long-term positive effect on worker's performance. My work adds task difficulty to this arsenal of behavioral tools that are costless to implement.

The rest of the chapter is organized as follows. Section 1.2 describes the experimental procedures, design and treatments. Section 1.3 presents the theoretical analysis and derives testable hypothesis. Section 1.4 discusses the results of the experiment and estimates a structural model motivated by the observed behavioral patterns. Section 1.5 concludes.

1.2 Experimental Design

1.2.1 Procedures

The experiment was conducted at the ExCEN lab at Georgia State University (GSU) in May–June 2015. A total of 98 subjects participated in the experiment over the course of six sessions. The subjects were recruited using the automated system that randomly picks the required number of participants from a pool of more than 2000 students who signed up for the participation in economic experiments. The subjects in the study were undergraduate students at GSU invited to participate via email. Upon arriving to the lab, the subjects reviewed and signed consent forms and took their seats in front of computers. Individual spaces had dividers on each side and at front to ensure privacy and independent decision-making. The subjects first read the experimental instructions at their own pace, and then an experimenter reviewed the instructions aloud while projecting the most important points on a screen. The subjects were allowed to ask questions at any point, and an experimenter answered them in private. Each session was run on computers, lasted for roughly 1.5 hours, and consisted of a math quiz, an effort task, a risk task and a demographic survey, presented in this order. The subjects received a show-up fee of $5 and the sum of the payoffs from the effort and risk decision tasks. They were paid privately in cash immediately after

the session. The average payment per subject was $45.89 with a standard deviation of $20.53 (the minimum payment was $5, and the maximum was $100), including the show-up fee.

Table 1.1 summarizes the demographic characteristics of the sample. It had a perfect gender balance with equal shares of males and females.[9] The racial composition was dominated by African American students: they account for 62% of the sample, while Caucasian students are just 14% of the sample. An average student's age was slightly above 21 years. The majority of subjects were in the advanced stages of a college program: more than half of participants were either juniors or seniors. Only 10% of the subjects came from an Economics or Finance major, which alleviates the concern that the observed behavior could be driven by the subjects sophisticated in economics.

Table 1.1: Socio-Demographic Characteristics of the Sample

Characteristic	Mean
Gender	
Male	0.48
Female	0.48
Race and Age	
White or Caucasian	0.14
Black or African American	0.62
Age	21.29
Year in Program and Major	
Freshman	0.04
Sophomore	0.18
Junior	0.24
Senior	0.45
Econ or Finance Major	0.10

[9]The shares do not sum up to one, since subjects had an option not to answer or state their gender as "other," though only 4% used one of these options.

1.2.2 Effort Task

In each round of the effort task,[10] a subject was given a project (see Figure 1.1) that had two possible outcomes: success or failure. Each project was characterized by the four characteristics (or treatment variables) that varied across rounds. These characteristics are: excess revenue (z), fixed revenue (w), difficulty (θ), and cost (k). In case of success the project yielded a high revenue (the sum of a fixed revenue plus an excess revenue), and in case of failure it yielded a low revenue (a fixed revenue only). The fixed revenue thus represents an unconditional (on performance) reward and the excess revenue represents a conditional reward. The difficulty of a project determined how the probability of success was computed. Higher difficulty resulted in a lower probability of success for any given level of effort; equivalently, in order to reach a given probability of success a subject had to choose higher effort in a more difficult project. The cost variable affected the steepness of the cost-of-effort function.

Subjects could choose an effort level, between 0 and 100 per cent, by moving a slider at the bottom of the screen.[11] The level of effort had a twofold effect: on the probability of success, and on a project's profit (revenue minus cost). Higher effort increased the probability of success but then led to the higher costs and consequently lower profits. The subjects incurred the cost of effort regardless of the outcome of a project. When deciding what effort level to choose, the subjects therefore faced a trade-off between a higher chance of the project being successful and higher costs. The subjects could observe how the values of probabilities, costs, and profits were changing as they experimented with the effort level. These variables were represented both numerically and graphically as colored bars.

The probability of success p as a function of effort a and difficulty θ was computed as

$$p(a,\theta) = a/2 + (1-\theta)/2, \tag{1.1}$$

[10] The task was programmed in z-Tree (Fischbacher, 2007).

[11] This experiment uses the chosen, rather than real, effort framework, as is common in tournament (Bull, Schotter, and Weigelt, 1987) and principal-agent experiments (Fehr, Gächter, and Kirchsteiger, 1997). The primary motivation for this was to allow for a clean test of theoretical predictions. Brüggen and Strobel (2007) demonstrate that the chosen effort framework yields qualitatively similar results to the real effort framework in their setting, while allowing for a greater control in the experiment.

so that, effectively, the probability of success was a simple average between the effort level and a project's "easiness," $1-\theta$.[12] The cost of effort was computed as the square of effort multiplied by the cost variable k:

$$c(a) = ka^2 \tag{1.2}$$

to induce a convex cost schedule. The subjects were informed of the linear relation between the probability of success and effort and the convex relation between the costs and effort.

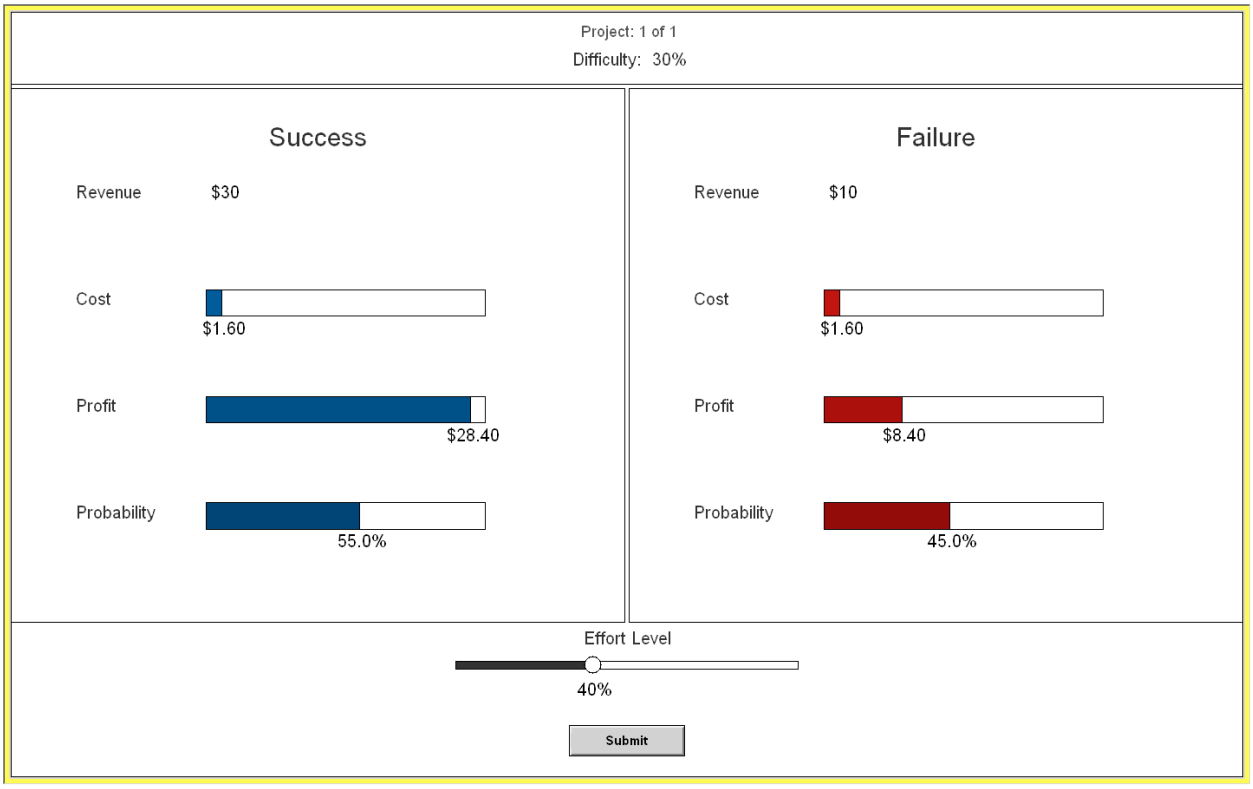


Figure 1.1: Choice Screen for the Effort Task

Once a subject clicked the "Submit" button, the outcome of the project was determined.[13] This was done graphically (see Figure 1.2) by presenting a bar with the success and failure regions, which corresponded to the success and failure probabilities determined by a subject's choice of

[12] In terms of a cognitive production function, this is equivalent to having constant returns to scale. This form was chosen to simplify the analysis and because it is easy to convey to subjects.

[13] An immediate feedback was provided after each round to ensure that subjects had a good understanding of the task, since quick feedback is crucial for learning and improving performance (Balzer, Doherty, and O'Connor, 1989; Hoch and Loewenstein, 1989). This was justified by the complexity of the decision task, which had many alternatives to choose from (the subjects could choose from 101 levels of effort) and several decision-relevant variables to consider.

effort. Along this bar, a white "needle" was moving quickly whose position at each fraction of a second was determined by a draw from a uniform random distribution. After three seconds, the needle stopped either in the success or failure region, which determined whether the project succeeded or failed.[14]

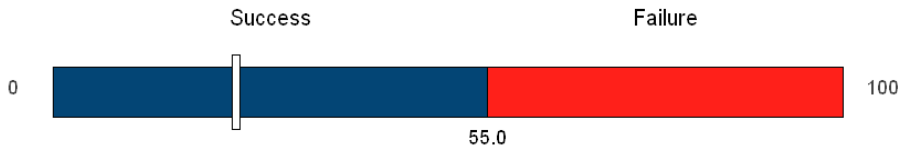


Figure 1.2: Project Outcome

After the outcome of the project was determined, the subjects could review the results of the current round on a summary screen, which presented the outcome of a project and the corresponding profit, along with some other details (see Figure 1.3). Each subject played between 15 and 19 rounds of the effort task,[15] which were preceded by the five practice rounds. After completing all the rounds, the subjects proceeded to a screen that determined the payoff for the task (see Figure 1.4). One round was randomly selected for payoff.[16] The screen showed all the rounds and profits made in those rounds. Every fraction of a second a random bar was highlighted, and after three seconds the highlighting stopped, which determined the payoff for the effort task.

The four project characteristics, difficulty, fixed revenue, excess revenue, and cost, changed across rounds, and the subjects were aware of this. Each distinct combination of the values of these treatment variables is a within-subjects treatment. The three monetary treatment variables had two possible values. The fixed revenue w assumed the values of 1 and 2, the excess revenue z assumed

[14]Realizing an outcome using a computer may raise credibility concerns among some subjects. This design was implemented due to its procedural convenience.

[15]The set of treatments varied slightly across sessions, see Appendix A.1 for details.

[16]While paying randomly for one round is theoretically not incentive compatible with the Rank Dependent Utility model (Harrison and Swarthout, 2014) that is used for subsequent analysis, the provision of feedback after each round (i.e., playing out choices sequentially but paying for one choice randomly in the end) might alleviate this concern in practice (but not in theory), as suggested by Cox, Sadiraj, and Schmidt (2015). They show that estimated risk preferences are not significantly different between a treatment in which subjects made, and were paid for, only one choice (theoretically incentive compatible with RDU) and a treatment in which subjects made multiple choices with feedback and were paid for a random round.

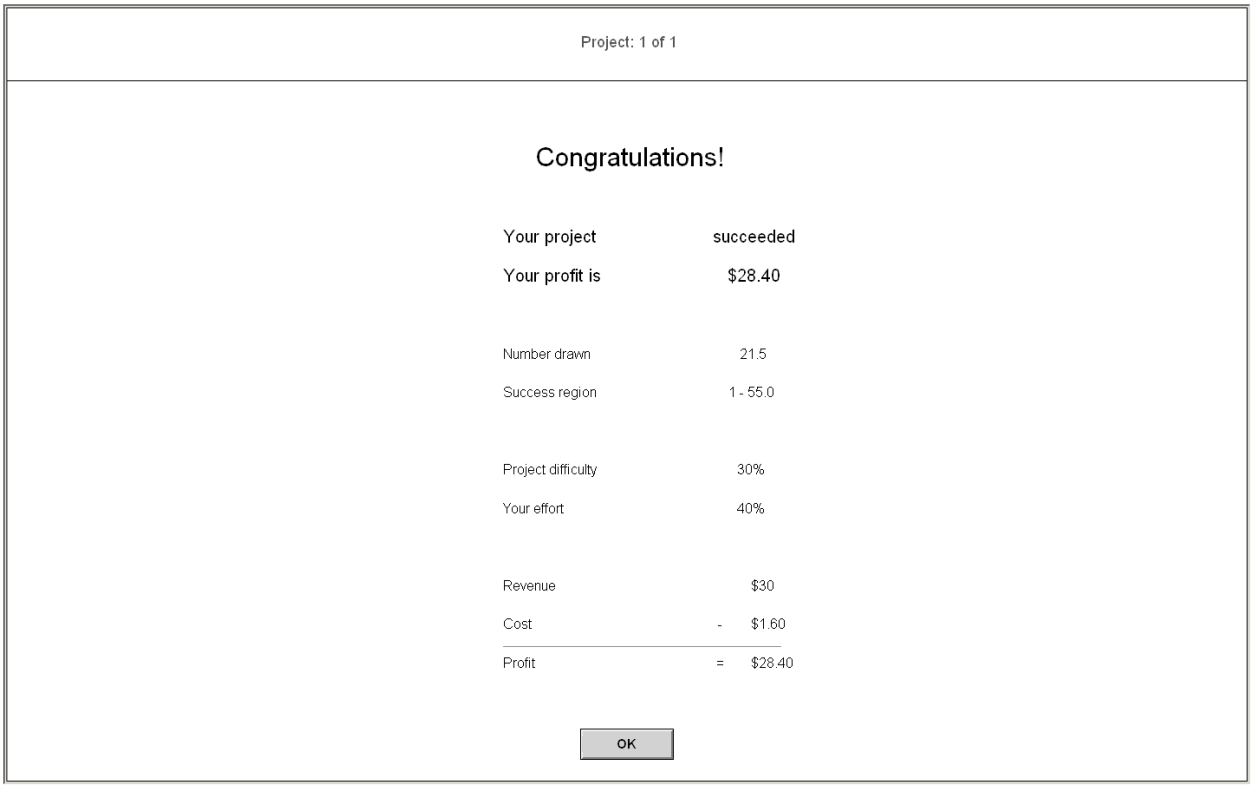


Figure 1.3: Project Summary

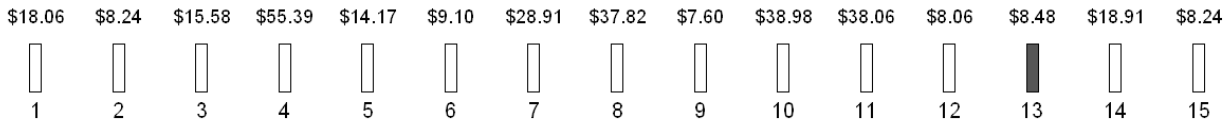


Figure 1.4: Payoff for the Effort Task

the values of 2 and 4, and the cost k assumed the values of 1 and 2. These values were multiplied by \$10 and then presented to the subjects, e.g., $w = 1$ and $z = 2$ corresponded to a project with a low revenue of \$10 and a high revenue of \$30 = \$10 + \$20. Similarly, given an effort level of 40% and $k = 2$, the cost of effort was \$3.2 = \$20$\times 0.4^2$. Difficulty assumed five values: 0, 0.25, 0.5, 0.75, and 1, which was important for identifying its potentially non-monotonic effect on effort. The treatments were constructed from the different permutations of the values of the treatment variables. The order of treatments was randomized for each subject. Appendix A.1 provides the summary of all treatments used in the experiment. The large set of different combinations of these

values allows one to study the effect of a given treatment variable across multiple combinations of the other three variables, which I exploit in the structural analysis below.

1.2.3 Lottery Task and Questionnaires

The lottery task presented the subjects with a sequence of binary lottery choices in the Hey and Orme (1994) format.[17] The lotteries were represented by pie charts with area sizes corresponding to outcome probabilities. The same information about outcomes and probabilities was also presented in a text format below the pie charts (see Figure 1.5). The subjects had to make a choice between a left and a right lottery. Appendix A.2 shows the battery of lotteries that were given to the subjects. The actual order and the position of lotteries on a screen (left or right) was randomized. The battery of lotteries used in the task is a variant of the one used by Harrison and Swarthout (2014), which is based on the lottery designs by Loomes and Sugden (1998) and Wakker *et al.* (1994). Together they provide a powerful test of EUT and RDU models.

After the subjects made all of their choices, one round was randomly selected for payoff. First, each subject rolled two dice that yielded a number between 1 and 100 that determined the paying round. Then the choice from that round was brought up on the screen, and the outcome was determined by the second roll of dice.

The math quiz (see Appendix A.3) consisted of six elementary questions on probabilities, expected value, and first-order stochastic dominance. They were designed to tune subjects into thinking about the concepts used in the experiment, as well as to provide additional controls for heterogeneity in choices. The demographic survey (see Appendix A.4) contained standard questions, such as age, gender, race, etc. The subject instructions for the two decision tasks can be found in Appendix A.5.

[17] The lottery task was programmed in Visual Basic.NET and is provided courtesy of J. Todd Swarthout.

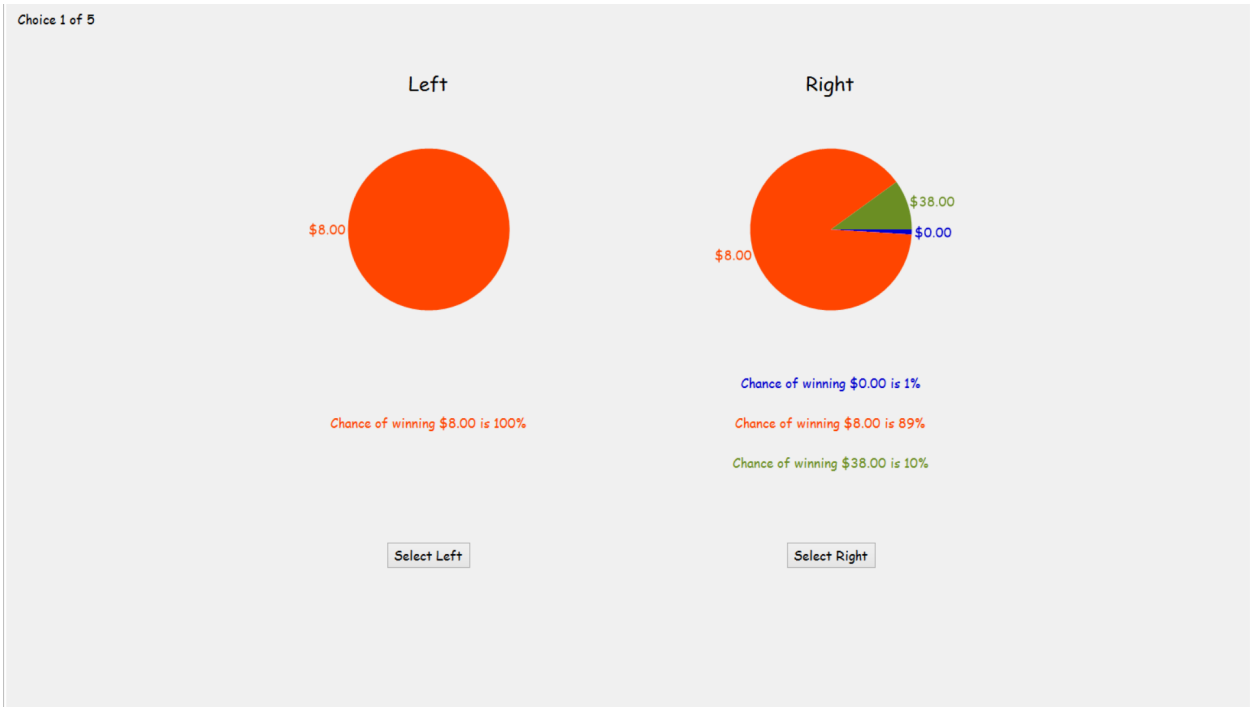


Figure 1.5: Lotteries Task

1.3 Theoretical Framework

1.3.1 General Model

This section presents a general model that does not make any parametric assumptions. The model highlights some issues with deriving clear-cut comparative statics predictions and how the experimental design overcomes them.

Consider an agent who works on a project with a binary outcome: success or failure. Success yields a high revenue, which is the sum of the fixed and excess revenue, $w+z$. Failure yields a low (fixed) revenue, w. It is assumed that $w,z \geqslant 0$. The outcome of the project is probabilistic, and the agent can improve the chances of success be exerting effort a, which is normalized to lie between 0 (no effort) and 1 (full effort). These chances are also affected by the difficulty θ of the project: a more difficult project is less likely to succeed for any given level of effort. The difficulty is normalized to be between 0 (the easiest project) and 1 (the hardest project). The vector of the values of a project characteristics is denoted as $\pi = (w,z,\theta)$.

I assume that the agent has preferences over money y and effort a represented by a utility function $u : \mathbb{R}_+ \times [0,1] \mapsto \mathbb{R}$. The u function is assumed to have the standard properties: it is twice continuously differentiable, strictly increasing and concave in money, and is strictly decreasing and concave in effort, i.e., the marginal disutility of effort increases.

The outcome of the project is a Bernoulli random variable X. The agent's revenue from the project is a random variable $Y = w + zX$. The probability of success function $p : [0,1] \times [0,1] \mapsto [0,1]$ depends on effort a and difficulty θ. I assume that p is twice continuously differentiable, increasing and concave in effort, and decreasing in difficulty. I do not make any assumptions about the concavity of p w.r.t. difficulty.

If the cdf of X conditional on effort and difficulty is $F(x \mid a, \theta)$ and pmf is $f(x \mid a, \theta) = xp(a,\theta) + (1-x)(1-p(a,\theta))$, then for $0 \leqslant x < 1$ we have $F_a(x \mid a, \theta) = -p_a(a,\theta) < 0$. This implies that the agent has an incentive to exert more effort, as the project endowed with a high effort level first-order stochastically dominates the project with a low effort level. However, there is a trade–off in that more effort leads to higher disutility from exerting it.

Since the project is risky, an assumption about the agent's risk preferences is required. I consider two alternative models and show that the nature of risk preferences plays a crucial role in determining the effect of a project's difficulty on the optimal effort choice. The first possible assumption about risk preferences is a standard EUT model.

Assumption 1.A (EU)**.** *The agent's risk preferences are characterized by the Expected Utility model:*

$$U(a \mid \pi) \equiv \mathbb{E}u(Y,a) = \sum_{x=0}^{1} u(w+zx,a) f(x \mid a, \theta).$$

I also explore the implications of the RDU model, which allows for non-linearity in probabilities. This property turns out to be critical for producing the non-monotonic effect of difficulty on effort.

Assumption 1.B (RDU)**.** *The agent's risk preferences are characterized by the Rank Dependent Utility model:*

$$\tilde{U}(a \mid \pi) \equiv \tilde{\mathbb{E}}u(Y,a) = \sum_{x=0}^{1} u(w+zx,a)\tilde{f}(x \mid a,\theta),$$

where $\tilde{f}(x \mid a,\theta) = x\tilde{p}(a,\theta) + (1-x)(1-\tilde{p}(a,\theta))$ *is the decision weight of an outcome x, and* $\tilde{p}(a,\theta) = \omega(p(a,\theta))$ *is the success probability weighted by the probability weighting function* $\omega : [0,1] \mapsto [0,1]$*, twice continuously differentiable and strictly increasing with* $\omega(0) = 0$ *and* $\omega(1) = 1$.

Under Assumption 1.A the agent chooses the optimal effort level a^* given the parameters of the problem π by maximizing $U(a \mid \pi)$. If $a^* = \arg\max U(a \mid \pi)$, then the first-order necessary condition must hold:

$$\mathbb{E}\left[u(Y,a^*)\frac{f_a(X|a^*,\theta)}{f(X|a^*,\theta)}\right] = -\mathbb{E}u_a(Y,a^*). \tag{1.3}$$

Equation (1.3) means that in the optimum the marginal benefit from exerting more effort on the left-hand side must be balanced by the marginal cost of effort on the right-hand side. The marginal benefit represents the expectation of the utility weighted by f_a/f, since the gain comes from the increased probability of success. It can be rewritten as $p_a\Delta u$, where $\Delta u \equiv u(w+z,a) - u(w,a)$ is the utility gain between the success and failure, given effort a. Written in this form, the marginal benefit is the marginal increase in the probability of success multiplied by the utility gain. The marginal cost is the expected marginal disutility of effort. It can also be rewritten using the Mean Value Theorem as $u_a + pzu_{ya}$, where the cross-partial is evaluated at some point $(\bar{y},a^*), \bar{y} \in [w, w+z]$. This form will be useful for understanding the comparative statics of the model.

In general, the first-order condition is not sufficient since the problem is not globally concave. The reason is that the second derivative of U, given by

$$U''(a \mid \pi) = \mathbb{E}u_{aa}(Y,a) + 2\mathbb{E}\left[u_a(Y,a)\frac{f_a(X|a,\theta)}{f(X|a,\theta)}\right] + \mathbb{E}\left[u(Y,a)\frac{f_{aa}(X|a,\theta)}{f(X|a,\theta)}\right]$$

is not always negative. The first term is the expectation of u_{aa}, which is negative by assumption. The third term can be rewritten as $\mathbb{E}[u_a f_{aa}/f] = p_{aa}\Delta u$, which is also negative, since the utility gain Δu is positive and p_{aa} is negative by assumption. The ambiguity comes from the second term, which equals $\mathbb{E}[u_a f_a/f] = p_a(u_a(w+z,a) - u_a(w,a))$. While $p_a > 0$ by assumption, the term in the brackets cannot be signed without an additional assumption about the cross-partial derivative of u. Using the Mean Value Theorem, the term in the brackets can be rewritten as $z u_{ya}(\bar{y},a)$, where $\bar{y}$ is some number on $[w, w+z]$. If u is submodular, $u_{ya} \leqslant 0$, the expected utility function U is strictly concave and the first-order condition is sufficient. If u is supermodular, $u_{ya} \geqslant 0$, however, one needs to check the second-order condition as well. In what follows, I assume that a^* is the interior global maximum of the agent's problem.[18] All the results still hold under Assumption 1.B as well, after replacing U, $\mathbb{E}$, f and p with $\tilde{U}$, $\tilde{\mathbb{E}}$, $\tilde{f}$ and $\tilde{p}$, respectively.

1.3.2 Special Cases

Before looking at the comparative statics results for the general case, I investigate several special cases of the utility function u that permit closed-form solutions.

The first case is an additively separable utility function, $u(y,a) = v(y) - c(a)$, which is the usual specification in the models of effort choice (Abeler, Falk, Goette, and Huffman, 2011; Hossain and List, 2012; Jayaraman, Ray, and de Vèricourt, 2016). Assume that $v : \mathbb{R}_+ \mapsto \mathbb{R}$ is twice continuously differentiable, increasing and concave. With a quadratic cost of effort as in (1.2) and a linear probability of success as in (1.1), the optimal effort is

$$a^* = \frac{v(w+z) - v(w)}{4k}.$$

A notable feature of this specification is that optimal effort does not depend on the project's difficulty. This is the consequence of the linearity of p and additive separability of u, under which the

[18] When U is not globally concave, it is possible to get a local maximum, in which case one also needs to check the function values at the boundaries of the choice set to find the global one, which will always exist by the extreme value theorem. Simulations show that for some parameter values, it is possible to run into corner solutions, which however does not invalidate the comparative statics results. They will still hold as weak inequalities.

cross-partial derivatives of both functions are zero. The optimal effort will increase with the excess revenue and decrease with the cost k, which is intuitive. The fixed revenue will cause the optimal effort to go down if v is strictly concave, which makes sense since if the agent can guarantee herself good result regardless of the outcome she will not have a strong incentive to exert effort. If v is linear, the optimal effort will not change with the fixed revenue.

The second case is a non-additively separable specification, in which effort has a monetary cost to the agent, just like in the current experimental design, and the utility of money is exponential, $u(y,a) = v(y-c(a)) = -e^{-\gamma(y-c(a))}$, with $\gamma > 0$ being the constant absolute risk aversion (CARA) parameter. The benefit of using the exponential (CARA) utility is its analytical tractability, which has been exploited in various settings, e.g. von Gaudecker, van Soest, and Wengstrom (2011). Assume again that the cost of effort is quadratic and the probability of success is linear. It can be shown that in this case the optimal effort is given by

$$a^* = \frac{A+\theta-\sqrt{(A+\theta)^2-2/(k\gamma)}}{2}, \text{ where } A \equiv \frac{1+e^{-\gamma z}}{1-e^{-\gamma z}},$$

and the effect of difficulty on the optimal effort is

$$\frac{da^*}{d\theta} = \frac{1}{2}\left(1-\frac{A+\theta}{\sqrt{(A+\theta)^2-2/(k\gamma)}}\right),$$

which is negative because of the positive risk aversion parameter. Unlike in the additively separable case, u has a positive cross-partial derivative, which drives this result. It is worth noting that the optimal effort will not depend on the fixed revenue w. As in the previous case, the optimal effort will increase with the excess revenue and decrease with the cost.

These two examples hint at the importance of the cross-partial derivatives of p and u in driving the effects of the projects' difficulty and monetary rewards. As will be shown shortly, the experimental design allows to control for both and thus yields clear-cut comparative statics predictions.

1.3.3 Comparative Statics

Turning now to the general case, I use the first-order condition (1.3) and the Implicit Function Theorem to perform a comparative statics analysis and show how the optimal effort level $a^*(\pi)$ varies with the treatment variables. Under Assumption 1.A, the effect of the project's difficulty on the optimal effort is given by

$$\frac{da^*(\pi)}{d\theta} = -\frac{z[p_\theta(a^*,\theta)u_{ya}(\bar{y},a^*) + u_y(\bar{\bar{y}},a^*)p_{a\theta}(a^*,\theta)]}{U''(a^* \mid \pi)}, \tag{1.4}$$

where $\bar{y}, \bar{\bar{y}}$ are some numbers on $[w, w+z]$. Note that the denominator is negative (a^* is the maximum) while z is positive. Since $p_\theta \leqslant 0$ and $u_y \geqslant 0$ by assumption, the sign of the effect depends on the signs of the cross-partial derivatives of u and p. Considering different possible cases, this effect can be concisely stated as follows.

Proposition 1.A. *If* $\operatorname{sgn}(u_{ya})\operatorname{sgn}(p_{a\theta}) < 1$, *then* $\operatorname{sgn}\left(\frac{da^*(\pi)}{d\theta}\right) = \operatorname{sgn}(\operatorname{sgn}(p_{a\theta}) - \operatorname{sgn}(u_{ya}))$.

The proposition says that in order to unambiguously sign the effect of difficulty on optimal effort, the cross-partial derivatives of u and p have to be either of the opposite signs, or one of them or both have to be zero. Then the effect will have the same sign as the sign of the cross-partial of p, if this derivative is not zero, or the opposite of the sign of the cross-partial of u, if this derivative is not zero. If both derivatives are zero, then the difficulty will have no effect on optimal effort, as we have seen in the additively separable case.

Assuming that $p_{a\theta} \neq 0$, the effect of difficulty will have the same sign as this cross-partial derivative. Intuitively, it implies that if effort and difficulty are complements in the probability function, it is optimal to increase effort in response to a higher difficulty. Since higher difficulty reduces the probability of success, the optimal response is to compensate this reduction. On the other hand, if effort and difficulty are substitutes in p, it is optimal to reduce effort, which makes the reduction in the probability of success even higher.

While it might not be immediately obvious why optimal effort would respond to a higher difficulty in such a manner, some insights might be gained by looking at the optimality condition

(1.3). Consider the case when $p_{a\theta} = 0$, for example, in which the effect will operate through a change in the marginal cost. If u is supermodular, Proposition 1.A predicts a decline in effort in response to higher difficulty. As the difficulty increases, the marginal benefit will not change since $p_{a\theta} = 0$. The marginal cost, however, will increase, since $u_{ya} \geqslant 0$ and an increase in difficulty leads to a decline in the probability of success. To restore the balance, the agent has to reduce effort. On the other hand, if u is submodular, the marginal cost will decrease as the difficulty goes up, and the agent will increase effort.

Using Assumption 1.B instead, the formula in (1.4) remains valid after the appropriate change of non-tilde characters to tilde characters. Expanding, one obtains

$$\frac{da^*(\pi)}{d\theta} = -\frac{z[\omega' p_\theta(a^*,\theta)u_{ya}(\bar{y},a^*) + u_y(\bar{\bar{y}},a^*)(p_{a\theta}(a^*,\theta)\omega' + p_\theta(a^*,\theta)p_a(a^*,\theta)\omega'')]}{\tilde{U}''(a^* \mid \pi)}, \quad (1.5)$$

where ω, ω' and ω'' are evaluated at $p(a^*,\theta)$. Note that if $\omega'' = 0$ we are back to the EU case, since there would be no probability weighting. The sign of the effect of difficulty on effort now depends, in addition to the signs of the cross-partial derivatives of u and p, on the sign of ω''. Assuming that $\omega'' \neq 0$, the effect of difficulty in this case can be concisely stated as follows.

Proposition 1.B. *If* $\operatorname{sgn}(u_{ya})\operatorname{sgn}(p_{a\theta}) < 1$ *and* $\operatorname{sgn}(u_{ya})\operatorname{sgn}(\omega'') > -1$, *then* $\operatorname{sgn}\left(da^*(\pi)/d\theta\right) = -\operatorname{sgn}(\omega'')$.

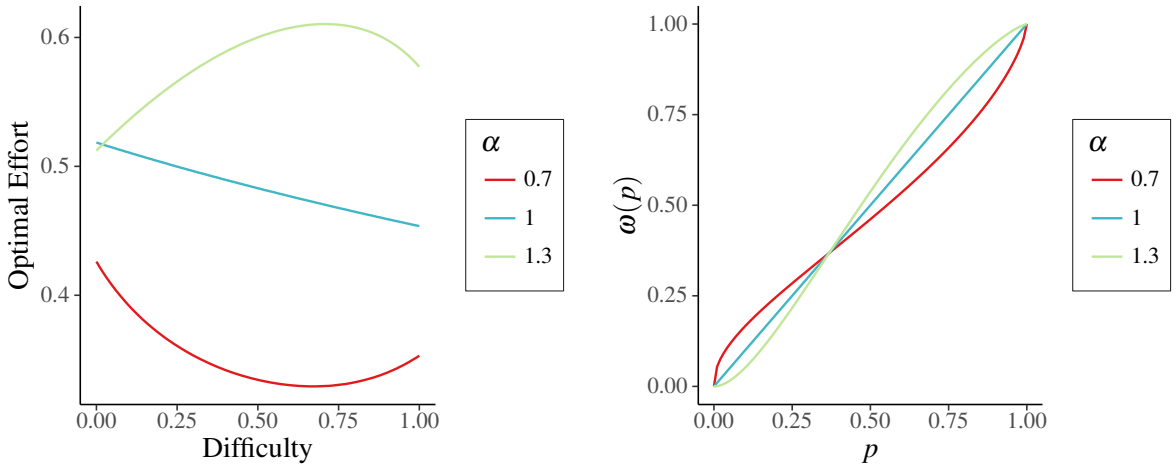
The proposition says that in order to sign the effect of difficulty unambiguously, the cross-partial derivatives of p and u have to be of the opposite signs as before (with the possibility of one or both of them being zero), while the cross-partial of u has be of the same sign as ω'' or equal to zero. If these two conditions hold, the effect of difficulty will go in the direction opposite to the sign of ω''. This result has an intuitive interpretation. If the agent exhibits probability pessimism, $\omega'' > 0$, she will reduce effort in response to a higher difficulty, as if she does not believe in her ability to affect the chances of success strong enough to "accept the challenge." On the other hand, if the agent exhibits probability optimism, $\omega'' < 0$, she will raise effort in response to a higher difficulty.

The propositions obtained so far predict only a monotonic effect of difficulty on effort. Given the findings in psychology on the inverse-U pattern of the response of effort to difficulty (Gendolla, Wright, and Richter, 2012; Gendolla, Richter, and Silvia, 2008), the question is whether either of the two models can produce a non-monotonic effect. For that, one would need to have a change in the sign of some terms as the difficulty changes. A natural candidate for this is ω'', since experiments sometimes find an inverse-S shape of the probability weighting function (Wu and Gonzalez, 1996; Bruhin, Fehr-Duda, and Epper, 2010). Inverse-S shaped weighting function is concave for low values of p and convex for high values of p.

Assume for simplicity that both cross-partial derivatives or u and p are zero and that ω is inverse-S-shaped. The sign of the effect of difficulty then would change from positive, for small p, to negative, for high p. The question is, however, how the effect of difficulty changes with θ itself and not with p. The difficulty is inversely related to the probability of success, and therefore one could expect that for high values of difficulty effort increases, while for low values of difficulty effort decreases, which implies a U-shaped pattern of response. The prediction is not clear-cut, since p and u might have non-zero cross-partial derivatives and the change of the sign of ω'' will violate the condition $\operatorname{sgn}(u_{ya})\operatorname{sgn}(\omega'') > -1$.

I resort to numerical simulations to understand the comparative statics in this case. Figure 1.6a shows the optimal effort as a function of difficulty for three different shapes of the probability weighting function. In the simulation, I use the monetary cost of effort specification with the CRRA utility of money $u(x) = x^{1-\gamma}/(1-\gamma)$ with $\gamma = 0.2$ and the one-parameter Prelec (1998) weighting function $\omega(p) = \exp(-(-\ln(p))^{\alpha})$. The cost and probability of success functions are specified as before.

As expected, optimal effort as a function of difficulty is U-shaped for an inverse-S-shaped probability weighting function (figure 1.6b shows the shape of ω for different values of α) and is inverse-U-shaped for an S-shaped probability weighting function. When there is no weighting, the optimal effort monotonically declines in difficulty as predicted by Proposition 1.A, since $u_{ya} > 0$ in this specification. These observations give rise to two possibilities. If one takes an inverse-S



(a) Effect of Difficulty on Effort

(b) Probability weighting function

Figure 1.6: Comparative Statics Under RDU

shape of ω as a starting point, then one should expect a U-shaped response of effort to difficulty. On the other hand, one might expect an inverse-U-shaped response of effort to difficulty, which would be the case of S-shaped probability weighting.

I now turn to the effects of monetary rewards on optimal effort. The effect of the excess revenue z is given by

$$\frac{da^*(\pi)}{dz} = -\frac{p(a^*,\theta)u_{ya}(w+z,a^*) + p_a(a^*,\theta)u_y(w+z,a^*)}{U''(a^* \mid \pi)}. \tag{1.6}$$

In the expression (1.6) above, the signs of all the terms are positive except for u_{ya} and the denominator, which is negative. One therefore obtains an unambiguous effect only in the case when u_{ya} is positive.

Proposition 2. *If $u_{ya} \geqslant 0$, optimal effort will increase with excess revenue.*

This prediction makes sense intuitively, as a higher excess revenue means a better outcome in case of success, which in turn justifies higher effort to improve the chances of success. Both the additively separable and exponential utility cases are examples of this. The result works, however, only when money and effort are complements in the utility function. To understand why it works this way, one should again look at the optimality condition (1.3). As the excess revenue goes up, the marginal benefit increases since u is increasing in money. The marginal cost decreases because

of the supermodularity of u. The equilibrium can be restored only by increasing effort, which pushes the marginal benefit down and the marginal cost up.

The effect of the fixed revenue on the optimal effort is given by

$$\frac{da^*(\pi)}{dw} = -\frac{\mathbb{E}u_{ya}(Y,a^*) + p_a(a^*,\theta)u_{yy}(\bar{y},a^*)\bar{y}}{U''(a^* \mid \pi)}.$$

The second term in the numerator is negative, since u is concave in money, and the denominator is negative. The effect of the fixed revenue is therefore unambiguous only when the sign of u_{ya} is also negative:

Proposition 3. *If $u_{ya} \leqslant 0$, optimal effort will decrease with fixed revenue.*

This means that if the agent can guarantee herself higher revenue regardless of the outcome, she has an incentive to become "lazy" and exert lower effort, provided that u is submodular. This is true, for example, in the additively separable case. To understand the intuition behind the result, note what happens to the optimality condition (1.3) as fixed revenue goes up. The marginal benefit decreases, since u is concave in money. The marginal cost, on the opposite, increases since the marginal utility of effort decreases with money by the submodularity assumption. The only way to restore the equilibrium between the marginal benefits and marginal costs is to reduce effort.

Propositions 2 and 3 will still hold for the RDU model, since $\tilde{p}$ and $\tilde{p}_a = \omega' p_a$ have the same signs as p and p_a, respectively.

1.3.4 Stochastic Choice

The model considered so far assumes that the agent's choices are deterministic, even though she makes them in a stochastic setting. For a fixed set of parameters $\bar{\pi}$, the agent always chooses $a^* = \arg\max U(a \mid \bar{\pi})$. However, a large body experimental of experimental evidence shows that subject's choices are stochastic (Starmer and Sugden, 1989; Camerer, 1989; Ballinger and Wilcox, 1997) and a large and growing theoretical literature rationalizes this behavior (Matêjka and McKay,

2015; Gul, Natenzon, and Pesendorfer, 2014). It is natural to ask whether and how the predictions change if one allows the agent's choices to be stochastic.

The main difference between a deterministic and a stochastic model of choice is that the stochastic model can only make predictions about *expected* effort, since choice is now a random variable for an outside observer. To illustrate the point, consider a simple case when the choice set consists of only two elements, full effort and no effort, $A = \{1, 0\}$. Assign probabilities q and $1 - q$ to choices of full and no effort, respectively, according to the usual multinomial logit formula (Luce, 1959)

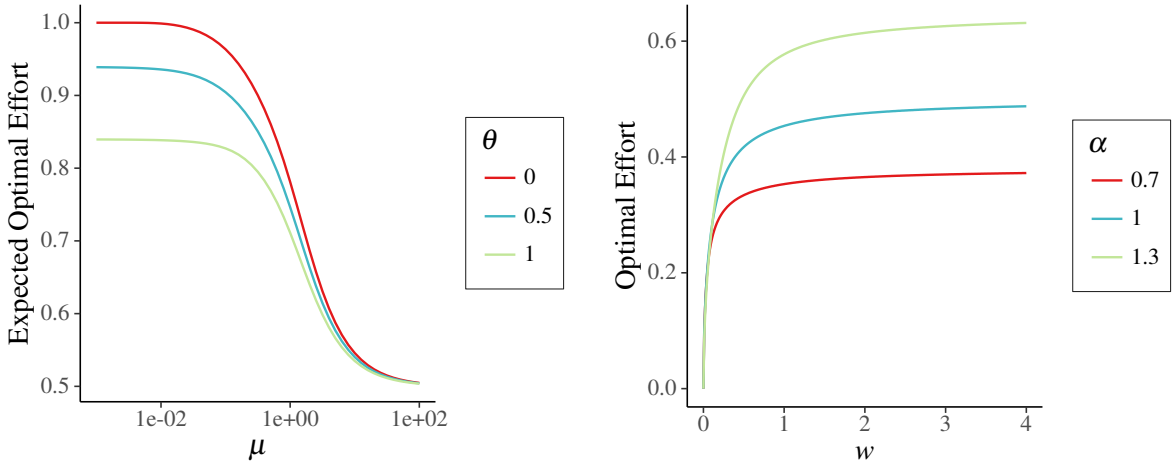
$$q = \frac{\exp(U(1 \mid \pi)/\mu)}{\exp(U(1 \mid \pi)/\mu) + \exp(U(0 \mid \pi)/\mu)} = \Lambda\left(\frac{U(1 \mid \pi) - U(0 \mid \pi)}{\mu}\right),$$

where Λ denotes the logistic cdf and μ is the noise parameter: higher values of μ make choices more "random" in the sense of not taking into account the respective utilities.

Suppose that under some set of parameters $\bar{\pi}$ the utility of full effort is higher than the utility of no effort, $\Delta U(\bar{\pi}) \equiv U(1 \mid \bar{\pi}) - U(0 \mid \bar{\pi}) > 0$. In this case, the deterministic model would predict that the choice will always be the full effort. In the stochastic model, we would have that the choices *on average* will be closer to the full effort, because $\mathbb{E}(a) = q$ and $\Delta U(\bar{\pi}) > 0$ implies that $p = \Lambda(\Delta U(\bar{\pi})/\mu) > 1/2$. In the limit, as noise goes to zero, one would have that $\lim_{\mu \to 0} \mathbb{E}(a) = 1$: small noise values would push the expected effort towards the prediction of the deterministic model, as the difference in utilities becomes more salient. As noise increases, the difference between utilities becomes less salient and the expected effort is sucked towards $1/2$, since $\lim_{\mu \to \infty} \Lambda(\Delta U(\bar{\pi})/\mu) = 1/2$: effort choice becomes uniformly distributed. The deterministic model is therefore a special case of a stochastic model with zero noise.

Figure 1.7a demonstrates the result in a general case when the agent's choice set is $[0, 1]$ for three levels of difficulty. I use the monetary cost of effort specification with the CRRA utility of money function $u(x) = x^{1-\gamma}/(1-\gamma)$ with $\gamma = 0.2$ and no probability weighting with the cost of effort and probability of success functions defined as before. Proposition 1.A predicts declining

effort levels in response to higher difficulty. The picture confirms that the ordering of optimal choices in the deterministic case is preserved for the expected choices in the stochastic case.



(a) Expected Effort as a Function of Noise

(b) Effect of Fixed Revenue on Effort

Figure 1.7: Effect of Noise and Fixed Revenue

1.3.5 Testable Hypotheses

In the experimental design, the probability of success function in (1.1) is linear, and therefore $p_{a\theta} = 0$. Moreover, the use of the chosen effort framework, rather than the real effort one, allows me to determine the sign of the cross-partial u_{ya} unambiguously. Since effort has a monetary cost to the agent, the utility function becomes $u(y,a) = v(y - c(a))$, where $v : \mathbb{R}_+ \mapsto \mathbb{R}$ is the utility of money, assumed to be twice continuously differentiable, increasing and concave, and $c : [0,1] \mapsto \mathbb{R}_+$ is the cost of effort function given by (1.2), twice continuously differentiable, increasing and convex. Then the cross-partial of the utility function is $u_{ya} = -c'v'' \geqslant 0$, which implies that $\operatorname{sgn}(da^*/d\theta) = -\operatorname{sgn}(u_{ya}) \leqslant 0$. Assuming the EU model 1.A, one can use Proposition 1.A to arrive at the following hypothesis.

Hypothesis 1.A (Difficulty, decreasing)**.** *Subjects' average effort will decrease with the project's difficulty.*

Under the assumption of RDU model 1.B and an inverse-S-shaped probability weighting function, Proposition 1.B together with numerical simulations imply an alternative hypothesis:

Hypothesis 1.B (Difficulty, U-shape)**.** *Subjects' average effort will first decrease, reach a minimum, and then increase with the project's difficulty.*

Assuming S-shaped probability weighting in the RDU model, one obtains the third alternative:

Hypothesis 1.C (Difficulty, inverse-U-shape)**.** *Subjects' average effort will first increase, reach a maximum, and then decrease with the project's difficulty.*

Turning to the effect of incentives, in the current design u is supermodular so that Proposition 2 applies directly leading to the following testable hypothesis:

Hypothesis 2 (Excess revenue)**.** *Subjects' average effort will increase with the project's excess revenue.*

In the experiment, $u_{ya} \geqslant 0$ and therefore Proposition 3 cannot be directly applied to sign the effect of fixed revenue. Note that in the special case of exponential utility the optimal effort level did not depend on the fixed revenue. An arguably more plausible case of CRRA utility does not have an analytical solution, but one can look at the numerical simulations to get some insights into the effect of fixed revenue on effort. Figure 1.7b shows the optimal effort as a function of w for $\gamma = 0.2$ and three parameter values of the probability weighting function. In all three cases effort monotonically increases. It is thus reasonable to expect the following behavior.

Hypothesis 3 (Fixed revenue)**.** *Subjects' average effort will either increase or not change with the project's fixed revenue.*

In the experiment, I vary one more project characteristic, k, which is the scaling factor of the cost of effort function $c(a) = ka^2$. Given the result for the exponential utility (which is also confirmed by numerical simulations for CRRA utility), one can expect the following behavior.

Hypothesis 4 (Cost of effort)**.** *Subjects' average effort will decrease with the cost of effort.*

1.4 Results

1.4.1 Aggregate-level Behavior

Unconditional effects

I begin analysis by looking at the pooled data and comparing the effort values across the different values of the treatment variables. Figure 1.8 presents the graphs for each of the treatment variables arranged in rows, with columns showing the means, histograms, and empirical CDFs of effort values. The first row shows the results for the excess revenue. On average, subjects choose higher effort values for a higher value of excess revenue. The mean effort for $z = 2$ is 0.532, while the mean effort for $z = 4$ is 0.642, a difference of 0.11 or 0.323 standard deviations. The difference between the mean effort values is highly significant (two-sided t-test, $p < 0.001$). Higher excess revenue is associated with a right shift of the distribution of effort values, as the middle figure shows, with the mass of choices shifting from 0, 0.5 and 0.75 to 1. The median effort for $z = 2$ is 0.51, while the median effort for $z = 4$ is 0.7. A Wilcoxon rank sum test shows that the effort values tend to be significantly higher for a higher value of z (two-sided test, $p < 0.001$). The CDFs graphs show a clear first-order stochastic dominance by the effort values for $z = 4$ relative to $z = 2$ and is confirmed by a Kolmogorov-Smirnov test (two-sided test, $p < 0.001$).

The effect of the fixed revenue on effort (the second row of Figure 1.8) is positive but less pronounced as compared to the effect of the excess revenue. The mean effort for $w = 1$ is 0.59, and the mean effort for $w = 2$ is 0.627, with a difference of 0.038 or 0.111 standard deviations.[19] The difference between the mean effort values is only marginally significant (two-sided t-test, $p =$ 0.056). While the higher value of the fixed revenue tends to shift the distribution of effort values to the right, it also leads to more choices of 0. The median effort values for $w = 1$ and $w = 2$ are 0.6 and 0.69, respectively. A Wilcoxon rank sum test shows that effort values tend to be higher for

[19] Whenever we perform a two-sample analysis of the effect of fixed revenue, we work with the subset of observations with $k = 1$. For the full sample, the fixed revenue and cost are highly correlated, which is due to the nature of the design. In the experiment, the cost cannot exceed the fixed revenue, otherwise a negative profit could occur and subjects could lose money. When analyzing the effect of cost, the sample is restricted to treatments with $w = 2$ for the same reasons.

a higher value of w, but the effect is, again, only marginally significant (two-sided test, $p = 0.048$). The CDF graph suggests a first-order stochastic dominance by effort values for $w = 2$, with a slight reversion of the effect for very small effort values. A Kolmogorov-Smirnov test, however, does not yield a significant result (two-sided test, $p = 0.088$).

The third row in Figure 1.8 shows that the cost has a clear negative effect on effort. The mean effort for $k = 1$ is 0.627, while the mean effort for $k = 2$ is 0.494, with a difference of -0.134 or -0.391 standard deviations. The difference between the mean effort values is highly significant (two-sided t-test, $p < 0.001$). Higher cost values tend to shift the distribution of effort values and the medians to the left. The median effort values for $k = 1$ and $k = 2$ are 0.69 and 0.5, respectively. A Wilcoxon rank sum test yields a highly significant result (two-sided test, $p < 0.001$). The CDF graph shows a clear first-order stochastic dominance by the effort values corresponding to $k = 1$, which is confirmed by a Kolmogorov-Smirnov test (two-sided, $p < 0.001$).

The last row of Figure 1.8 reveals a non-monotonic inverse-U effect of difficulty on effort. The mean effort values for $\theta = 0, 0.25, 0.5, 0.75$ and 1 are 0.573, 0.656, 0.657, 0.601, and 0.54, respectively. The difference between the peak mean effort at $\theta = 0.5$ and mean effort at $\theta = 0$ is 0.083 (two-sided t-test, $p = 0.002$) or 0.244 standard deviations, which is twice as high as the effect of fixed revenue and slightly lower than the effect of excess revenue. Increasing difficulty from $\theta = 0.5$ to $\theta = 1$ reduces the mean effort by 0.117 (two-sided t-test, $p < 0.001$) or 0.341 standard deviations, which exceeds the effect of excess revenue. The difference between the mean effort values corresponding to the lowest and the highest values of difficulty is 0.033, however, this diffirence is not statistically significant (two-sided t-test, $p = 0.103$). The distribution of efforts tends to shift to the right as difficulty increases from 0 to 0.5 and then back left as the difficulty continues to increase. The median effort values for the five consecutive values of difficulty are 0.555, 0.69, 0.7, 0.695, and 0.5. A Wilcoxon rank sum test shows a significant increase in the values of effort corresponding to $\theta = 0.5$ relative to $\theta = 0$ (two-sided test, $p = 0.003$), and a significant decrease in the values of effort corresponding to $\theta = 1$ relative to $\theta = 0.5$ (two-sided test, $p < 0.001$). The CDF graphs corresponding to the extreme (0 and 1) and the middle (0.5)

levels of difficulty, confirm the inverse-U pattern except for the high values of effort, when the CDF corresponding to $\theta = 1$ shifts to the right relative to the other two CDFs. A Kolmogorov-Smirnov test confirms the first-order stochastic dominance of effort values corresponding to $\theta = 0.5$ relative to both effort values corresponding to $\theta = 0$ (two-sided test, $p = 0.004$) and effort values corresponding to $\theta = 1$ (two-sided test, $p < 0.001$).

The pooled analysis suggests that effort responds positively to higher excess revenue, higher fixed revenue, and lower cost. Effort responds positively to the initial increase in difficulty and negatively to the further increase in difficulty. I perform two additional analyses to see whether these results are robust. First, I make use of the within-subject design and conduct paired-samples tests. For each subject, I take average effort for the two values of each treatment variable (the change in difficulty is broken down in two steps: from 0 to 0.5 and from 0.5 to 1). I use three paired-samples tests, the paired t-test, the Wilcoxon signed rank test, and the sign test, to see whether there are significant differences in effort between the two samples. Table 1.2 shows the results. They confirm the patterns revealed by the pooled analysis.

Table 1.2: Summary Results for Paired-Samples Tests

		p-values		
Treatment Variable	ATE	t	Wilcoxon	sign
z	0.107	<0.001	<0.001	<0.001
w	0.032	0.069	0.04	0.025
k	−0.124	<0.001	<0.001	<0.001
θ (0 to 0.5)	0.077	0.012	0.001	<0.001
θ (0.5 to 1)	−0.110	0.001	0.002	0.01

Notes: The first column shows the average treatment effects of treatment variables. The last three columns show the p-values from the paired t-test, the Wilcoxon signed rank test, and the sign test. All the tests are two-sided.

The experimental design allows for a second robustness check based on using *ceteris paribus* pairs. A pair of samples of effort values is called a *ceteris paribus* pair (*CP*-pair) for a treatment variable x, if only x changes within a pair while the rest of the treatment variables are kept constant. Let δ denote the vector of the values of treatment variables. Each treatment $j \in J$ represents a

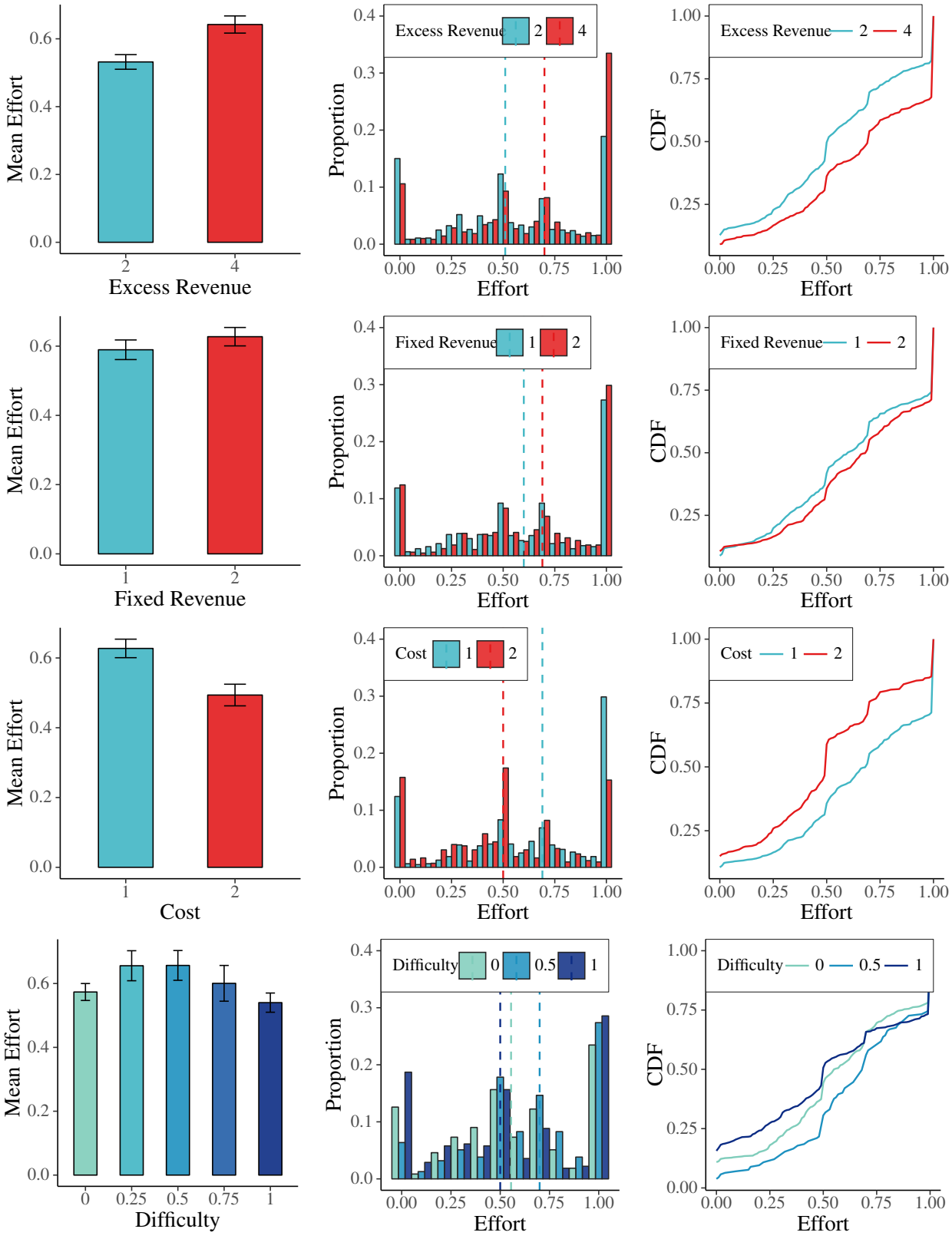


Figure 1.8: Average Treatment Effects, Histograms, and CDFs for Treatment Variables

particular combination of the values of treatment variables from the set of all treatments J, $\delta_j \equiv (z_j, w_j, k_j, \theta_j)$. The set J_x is the set of all *CP*-pairs for a treatment variable x and is defined as

$$J_x = \left\{ (j_1, j_2) \mid j_1, j_2 \in J, x_{j_1} = x^1, x_{j_2} = x^2, x^1 < x^2, (\delta_{-x})_{j_1} = (\delta_{-x})_{j_2} \right\},$$

where δ_{-x} denotes the vector of the values of treatment variables excluding x, and for convenience the first index corresponds to a treatment with a lower value of the treatment variable. Let k index the elements of J_x and $K_x \equiv ||J_x||$ be the number of elements in this set. For every *CP*-pair $k = 1, \ldots, K_x$ the two samples used in the *ceteris paribus* tests are

$$\left\{ a_{i,(j_1)_k} \right\}_{i=1,\ldots,n}, \left\{ a_{i,(j_2)_k} \right\}_{i=1,\ldots,n}, (j_1, j_2)_k \in J_x,$$

where $a_{i,j}$ denotes effort chosen by a subject i in a treatment j, and n is the total number of subjects in the experiment. The *ceteris paribus* test is a strict test of hypotheses, since they are derived from the comparative statics results, which assume that only one independent variable changes while others are held constant.

Table 1.3 shows the results. The effect of excess revenue is positive and highly significant in almost all *CP*-pairs and across all the three tests. The effect of fixed revenue changes sign across the *CP*-pairs and is not significant in any of them. The effect of the cost is negative in all *CP*-pairs and significant in most of them. The effect of increasing difficulty from 0 to 0.5 is positive in all the *CP*-pairs and significant in half of them. The effect of the further increase in difficulty is negative and significant in all the *CP*-pairs but one. The mixed results across different pairs call for some way of aggregating them in a single number.

To aggregate the results across multiple *CP*-pairs, I use the following empirical model:

$$a_{ij} = \beta_1 + \beta_2 \mathbb{I}(\delta_{ij} = \delta_2) + \ldots + \beta_J \mathbb{I}(\delta_{ij} = \delta_J) + \varepsilon_{ij}, \tag{1.7}$$

Table 1.3: Summary Results for the *Ceteris Paribus* Analysis

		p-values		
CP-pair	ATE	*t*	Wilcoxon	sign
		z		
1	0.083	0.003	<0.001	<0.001
2	0.05	0.153	0.042	0.007
3	0.146	<0.001	<0.001	<0.001
4	−0.002	0.971	0.765	0.856
5	0.174	0.004	0.001	
6	0.174	0.016	0.006	
7	0.167	<0.001	0.001	0.014
8	0.202	<0.001	<0.001	0.001
9	0.154	<0.001	<0.001	0.001
		w		
1	0.011	0.7	0.463	1
2	−0.022	0.407	0.598	0.902
3	0.077	0.187	0.245	
4	−0.027	0.634	0.558	
5	0.035	0.617	0.714	
6	0.015	0.688	0.835	0.815
7	0.05	0.304	0.25	0.228
		k		
1	−0.173	<0.001	<0.001	<0.001
2	−0.077	0.019	0.001	<0.001
3	−0.15	0.024	0.089	
4	−0.007	0.93	0.654	
5	−0.01	0.926	0.876	
6	−0.073	0.036	0.112	0.336
7	−0.121	0.002	0.002	<0.001
		θ (0 to 0.5)		
1	0.055	0.292	0.25	
2	0.018	0.729	0.816	
3	0.183	0.039	0.042	
4	0.142	0.015	0.038	
		θ (0.5 to 1)		
1	−0.165	0.004	0.004	
2	−0.123	0.023	0.028	
3	−0.189	0.036	0.045	
4	−0.096	0.105	0.394	

Notes: The first column is an index of a *CP*-pair, the second column shows the average treatment effect of a treatment variable in a given pair, the last three columns show the *p*-values from either a) the paired *t* test, the Wilcoxon signed rank test, and the sign test, or b) the unpaired *t* test and the Wilcoxon rank sum test. The *CP*-pairs with an empty value in the column for the sign test are the non-paired samples. All the tests are two-sided.

where $\mathbb{I}(x)$ is the indicator function. The model represents a non-parametric regression of effort on all possible treatments. The regression uses robust standard errors clustered at the subject level to account for a within-subject correlation of responses. Define $\beta_j^* = \beta_j + \beta_1$ for $j = 2,\ldots,J$ and $\beta_1^* = \beta_1$. Then for some *CP*-pair k, the estimated average treatment effort of increasing treatment variable x from x^1 to x^2 is

$$\begin{aligned}\text{ATE(CP)}_x^k &= \mathbb{E}[a \mid x = x^2, \delta_{-x} = (\delta_{-x})_k] - \mathbb{E}[a \mid x = x^1, \delta_{-x} = (\delta_{-x})_k] \\ &= \hat{\beta}^*_{(j_2)_k} - \hat{\beta}^*_{(j_1)_k},\end{aligned}$$

where $(j_1, j_2)_k \in J_x$. The aggregate effect across all the *CP*-pairs can be computed as an average of the estimated effects in each *CP*-pair

$$\text{ATE(CP)}_x = \frac{1}{K_x}\left(\hat{\beta}^*_{(j_2)_1} + \ldots + \hat{\beta}^*_{(j_2)_{K_x}} - \hat{\beta}^*_{(j_2)_1} - \ldots - \hat{\beta}^*_{(j_2)_{K_x}}\right). \tag{1.8}$$

The statistical significance of the aggregated effect can be tested as a restriction that the linear combination of the coefficients given by (1.8) is zero using Wald test. Table 1.4 shows the results. It also contains the average treatment effects for the full samples not restricted to *CP*-pairs. These effects are computed from the empirical model simply as

$$\text{ATE}_x = \mathbb{E}[a \mid x = x^2] - \mathbb{E}[a \mid x = x^1] = \frac{1}{||L_x^2||}\sum_{l_2 \in L_x^2} \hat{\beta}^*_{l_2} - \frac{1}{||L_x^1||}\sum_{l_1 \in L_x^1} \hat{\beta}^*_{l_1},$$

where $L_x^1 = \{l_1 \in J \mid x_{l_1} = x^1\}, L_x^2 = \{l_2 \in J \mid x_{l_2} = x^2\}$, $x^1 < x^2$, and $\beta_j^* = \beta_j + \beta_1$ for $j = 2,\ldots,J$ and $\beta_1^* = \beta_1$.

The table confirms the previous patterns, however, the positive effect of fixed revenue turns out to be insignificant. Overall, the analysis shows that the positive effect of excess revenue, the negative effect of cost, and the inverse-U effect of difficulty are highly robust. The effect of fixed revenue, on the other hand, is weak and disappears in the regression model. The following result summarizes the effects of the treatment variables and their relation to the hypotheses.

Table 1.4: Aggregate Results

	CP-pairs		Full samples	
Treatment variable	ATE	*p*-value	ATE	*p*-value
z	0.128	<0.001	0.109	<0.001
w	0.020	0.369	0.007	0.772
k	−0.086	0.001	−0.102	<0.001
θ (0 to 0.5)	0.099	0.003	0.081	0.016
θ (0.5 to 1)	−0.129	<0.001	−0.114	0.001

Notes: The second and third columns show the average treatment effect of increasing a treatment variable across all the *CP*-pairs and the *p*-value of a Wald test. The last two columns show the same quantities computed using the full samples that are not restricted to *CP*-pairs.

Result 1. *The effects of the treatment variables are as follows:*

1. *The initial increase in difficulty results in an increase in subjects' effort. After reaching a peak at $\theta = 0.5$, effort declines with the further increase in difficulty. The result is consistent with Hypothesis 1.C but not 1.A or 1.B.*

2. *The increase in excess revenue results in an increase in subjects' effort, in accordance with Hypothesis 2.*

3. *The increase in fixed revenue results in a weak increase in subjects' effort, in accordance with Hypothesis 3.*

4. *The increase in the cost of effort results in a decrease in subjects' effort, in accordance with Hypothesis 4.*

The positive effect of excess revenue is in line with the previous findings in the literature on the strong positive effect of conditional rewards (Hossain and List, 2012; Lazear, 2000). The literature suggests that when conditional rewards are too small or too high, they can backfire and lead to lower effort either through crowding out of intrinsic motivation or chocking-under-pressure (Gneezy, Meier, and Rey-Biel, 2011). I do not observe these negative effects in my data because the chosen effort framework leaves little space for these two channels and also because the level of

conditional rewards in the experiment apparently did not hit either extreme. The weakly positive effect of fixed revenue is also consistent with the previous studies on the effect of unconditional rewards. A common finding is that unconditional rewards are effective in the short-run (Gneezy and List, 2006; Jayaraman, Ray, and de Vèricourt, 2016) and when the reciprocity channel exists (Cohn *et al.*, 2014; Hennig-Schmidt *et al.*, 2010). Since the subjects in the present experiment were unlikely to have reciprocal motives in the task—their performance did not benefit anyone else but them—the increase in the fixed revenue did not result in a significant increase in effort. The effort task is static in nature and thus one cannot observe any dynamic effects of unconditional rewards. The inverse-U effect of difficulty is consistent with the findings in psychology (Richter *et al.*, 2008). The RDU model can rationalize the inverse-U pattern if the probability weighting function is S-shaped, which is the opposite of what is sometimes reported.[20] I discuss this point in more detail in the structural analysis section.

Interaction effects

This section looks at the *interacted* average treatment effects, which reveal the complementarity or substitutability patterns between the treatment variables. Define the ATE of increasing a treatment variable x from x^1 to x^2, conditional on a variable y taking a value of y^k as

$$\text{ATE}_x(y = y^k) \equiv \mathbb{E}[a \mid x = x^2, y = y^k] - \mathbb{E}[a \mid x = x^1, y = y^k].$$

The interaction effect of increasing x from x^1 to x^2 conditional on $y = y^2$ relative to $y = y^1$, $y^1 < y^2$ is then given by

$$\text{ATE}_{xy} \equiv \text{ATE}_x(y = y^2) - \text{ATE}_x(y = y^1).$$

A positive interaction effect between variables x and y means that the ATE of x is higher for a higher value of y. When both treatment variables x and y have positive ATE's, the interpretation is that the increase in one of the variables amplifies the effect of the other. On the other hand, when

[20] Wilcox (2015b) finds little support for the commonly used inverse-S specification. Most of the subjects in that study uniformly overweight best outcomes.

one of the treatment variables, say x, has a positive ATE while the other one, say y, has a negative ATE, the interpretation of $\text{ATE}_{xy} > 0$ is that y amplifies the effect of x, but x dampens the (negative) effect of y (by reducing the absolute value of the negative effect). If $\text{ATE}_{xy} < 0$, the two variables are substitutes, and the interpretation is similar. Clearly, the matrix of the interaction effects is symmetric, $\text{ATE}_{xy} = \text{ATE}_{yx}$.

Table 1.5 presents this matrix, computed from the regression (1.7). The first row shows that excess and fixed revenues turn out to be complements: both of them have positive ATEs, and thus they amplify the effect of each other. Cost tends to dampen the effect of excess revenue, since $\text{ATE}_{zk} < 0$. The effect of the initial (from 0 to 0.5) increase in difficulty amplifies the effect of excess revenue, but the further (from 0.5 to 1) increase in difficulty has no significant impact on the ATE of excess revenue. This means that excess revenue is most effective in stimulating effort when difficulty is intermediate to high. This result is consistent with Vandegrift and Brown (2003) who find that conditional rewards have little effect when the task is easy but become effective for more difficult tasks.

The second row in Table 1.5 shows the results for fixed revenue.[21] The initial increase in difficulty has a positive impact on the ATE of fixed revenue, while the further increase has a negative impact. The effect of fixed revenue is thus most pronounced at the intermediate level of difficulty. The third row in Table 1.5 shows the results for cost. As noted above, cost and excess revenue are substitutes, so that excess revenue amplifies the negative effect of cost. The initial increase in difficulty dampens the negative effect of cost, while the further increase in difficulty has no significant effect. The findings about the interaction effects are summarized below.

Result 2. *Excess and fixed revenue have a complementary effect on effort, amplifying the positive effect of each other. Excess revenue and cost are substitutes. The intermediate level of difficulty makes the positive effect of excess and fixed revenue most pronounced, while dampening the negative effect of cost.*

[21] I cannot compute ATE_{wk} (or ATE_{kw}) because cost cannot exceed fixed revenue in the experiment, and the quantity $\mathbb{E}[a \mid w = 1, k = 2]$ needed to compute the interaction effect is not observed.

Table 1.5: Matrix of Interaction Effects

	z	w	k	θ (0 to 0.5)	θ (0.5 to 1)
z		0.039	−0.057	0.075	0.006
		(<0.001)	(<0.001)	(<0.001)	(0.356)
w	0.039			0.097	−0.044
	(<0.001)			(<0.001)	(<0.001)
k	−0.057			0.053	−0.006
	(<0.001)			(<0.001)	(0.5)
θ (0 to 0.5)	0.075	0.097	0.053		
	(<0.001)	(<0.001)	(<0.001)		
θ (0.5 to 1)	0.006	−0.044	−0.006		
	(0.356)	(<0.001)	(0.5)		

Notes: Each cell in the table shows the change in the ATE of a row variable when a column variable increases. The effect of difficulty is broken down into the initial increase (0 to 0.5) and the further increase (0.5 to 1). The p-values of a test of a null hypothesis of a zero effect are in parenthesis.

1.4.2 Individual-level Heterogeneity

This section explores the individual heterogeneity in the subjects' responses to the treatment variables. Figure 1.9 shows the distributions of the average treatment effects for each treatment variable computed at the subject level. The coloring highlights the probability masses of the negative (red) and positive (blue) ATE's. The subjects are typed based on the sign of their ATE as either Decreasing (negative ATE) or Increasing (positive ATE) for excess revenue, fixed revenue, and cost. The typing for difficulty is more complicated and explained below.

All the five distributions in Figure 1.9 are roughly bell-shaped and symmetric. The ATE's for each treatment variable vary across subjects both in terms of magnitudes and signs. The distribution of the ATE's for the excess revenue is unlikely to be normal (Shapiro-Wilk test, $p = 0.02$). The majority of subjects, 81%, increase their effort in response to higher excess revenue. The proportion of Increasing types is significantly higher than the proportion of Decreasing types (test for equality of proportions, $p < 0.001$). The behavior of Decreasing types is hard to rationalize (it would imply that those subjects prefer less money to more money), and thus is probably due to confusion. While the proportion of Decreasing types is non-negligible, the mean ATE for them

is only -0.059. Excluding the subjects who make errors would increase the mean effect size to 0.147 from the unconditional average of 0.107.

The ATE's for the fixed revenue are well approximated by a normal distribution (Shapiro-Wilk test, $p = 0.621$). The increase in the fixed revenue results in higher effort for 62% of the subjects. The remaining 38% reduce their effort on average by 0.14, which is comparable to the mean effect size for Increasing types, 0.136. The proportion of Increasing types is significantly higher than the proportion of Decreasing types (test for equality of proportions, $p = 0.02$), but the mean effect size across all the subjects, 0.032, is close to zero. The behavior of Decreasing types, given their strong presence in the distribution, is unlikely to be entirely driven by errors and might reflect actual preferences.

The ATE's for the cost are also well approximated by a normal distribution (Shapiro-Wilk test, $p = 0.794$). The effect of increasing cost is negative for 74% of the subjects. The proportion of Decreasing types is significantly higher than the proportion of Increasing types (test for equality of proportions, $p < 0.001$). The proportion of Increasing types is non-negligible, but the mean effect size for them, 0.087, is relatively small. Their behavior is likely to be caused by errors. Excluding them would reduce (increase in absolute terms) the mean effect size from -0.124 to -0.196.

The initial increase in difficulty (from 0 to 0.5) results in higher effort for 71% of the subjects. The mean effect size for those subjects is 0.219. The distribution of the ATE's for the initial increase in difficulty is unlikely to be normal (Shapiro-Wilk test, $p = 0.001$). The increase in difficulty from 0.5 to 1 leads to lower effort for 62% of the subjects, with the conditional mean effect size of -0.308. The ATE's for the further increase in difficulty are well approximated by a normal distribution (Shapiro-Wilk test, $p = 0.968$).

The subjects who increase their effort initially are not necessarily the same subjects who then reduce their effort (e.g., some of them increase their effort even further). The response to difficulty is thus characterized by Increasing, Decreasing, U, or Inverse-U types,[22] which requires knowing

[22]Specifically, Increasing types increase their effort on both intervals of difficulty (from 0 to 0.5 and 0.5 to 1), Decreasing types reduce their effort on both intervals, U types reduce their effort on the first interval and increase their effort on the second interval, and Inverse-U types increase their effort on the first interval and reduce their effort on the second interval. I omit the intermediate values of difficulty 0.25 and 0.75 to simplify the classification of types.

the full response path and cannot be shown on a distributions picture like in Figure 1.9. The dominant type is Inverse-U, with 50% of subjects conforming to it (see Table 1.6d). The equality of proportions of each type is clearly rejected (test for equality of proportions, $p < 0.001$). The subjects are split roughly equally among the three other types, with Increasing type being the second largest group after Inverse-U type. While the theory permits both Decreasing and U-shaped patterns of responses, these patterns do not necessarily represent actual preferences, since their prevalence is of the same magnitude as the proportion of apparently erratic responses to the excess revenue and cost.

Result 3. *Subjects vary both in terms of the magnitudes and signs of their ATE's. The distributions of the ATE's are roughly bell-shaped, symmetric, and likely to be normal in three out five cases. One-fifth to one-quarter of the subjects are likely to make errors in their responses, but the average size of the errors is small.*

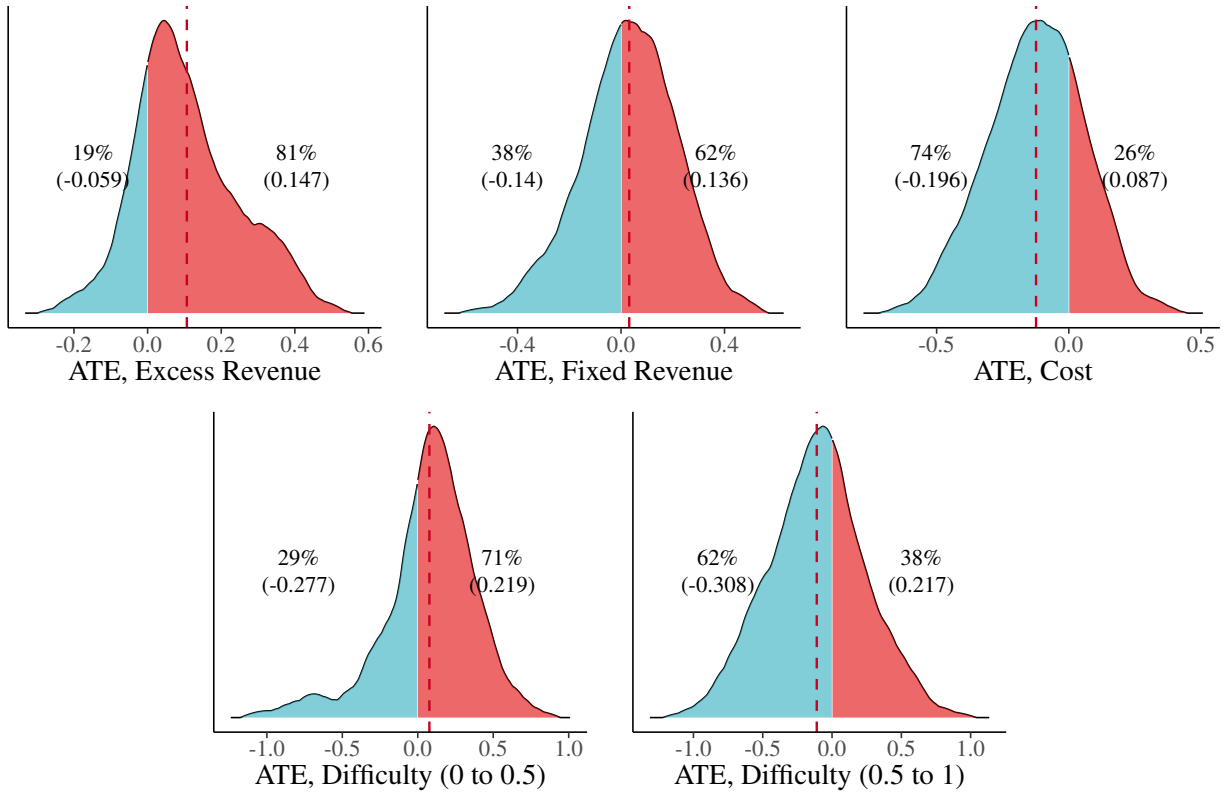


Figure 1.9: Kernel Density Plots of the Average Treatment Effects by Subject

An important question is whether the differences between the types can be attributed to the observable subjects' characteristics, such as gender. In the cases of excess revenue and cost, this amounts to asking whether females or males are more likely to make errors. Table 1.6 shows the contingency tables for each treatment variable, with subjects' gender in rows and subjects' response type in columns. For the excess revenue (Table 1.6a), the proportion of females belonging to a decreasing type, 17%, is lower than the corresponding proportion of males, 23%, however, this difference is not statistically significant (Fisher's exact test, $p = 0.608$). A similar result holds for the fixed revenue (Table 1.6b), the proportion of females belonging to a decreasing type, 34%, is slightly lower than the corresponding proportion of males, 40%, but not significantly so (Fisher's exact test, $p = 0.67$). The proportion of females who decrease their effort in response to higher cost, 70%, (Table 1.6c) is slightly lower than the corresponding proportion of males, 79%, with no significant difference (Fisher's exact test, $p = 0.478$). The effect of difficulty (Table 1.6d), however, does reveal a marginally significant association between gender and response type (Fisher's exact test, $p = 0.048$). The proportion of males belonging to the Inverse-U type, 64%, is higher than the corresponding proportion of females, 36%. The females are more likely than males to exhibit the U and Increasing effort response to difficulty.

Result 4. *The data do not indicate significant differences between genders in terms of their propensity to make errors, or respond to fixed revenue. The inverse-U response to difficulty is more likely to be observed among males than among females, while females are more likely to exhibit the U and Increasing response.*

1.4.3 Structural Analysis

In this section I ask whether a simple structural model can explain the observed behavioral patterns. As highlighted by the theoretical analysis in Section 1.3, the inverse-U pattern of effort response to difficulty can be accommodated by the RDU model, but not by the EUT model. Therefore, I will be estimating the parameters of the RDU model.

Table 1.6: Association Between Gender and Response Type by Treatment Variable

(a) Excess revenue

	Decreasing	Increasing	Total
Female	8	39	47
Male	11	36	47
Total	19	75	94

Fisher's exact test, $p = 0.608$

(b) Fixed revenue

	Decreasing	Increasing	Total
Female	16	31	47
Male	19	28	47
Total	35	59	94

Fisher's exact test, $p = 0.67$

(c) Cost

	Decreasing	Increasing	Total
Female	33	14	47
Male	37	10	47
Total	70	24	94

Fisher's exact test, $p = 0.478$

(d) Difficulty

	U	Inv-U	Decr	Incr	Total
Female	10	17	7	13	47
Male	5	30	6	6	47
Total	15	47	13	19	94

Fisher's exact test, $p = 0.048$

Consider an agent with preferences β who faces treatment δ. The rank-dependent utility of effort choice $a \in A = \{0, 0.01, \ldots, 1\}$ will be given by

$$U(a \mid \delta, \beta) = \omega(p(a, \theta) \mid \beta_p) u(w + z, a \mid \beta_u) + (1 - \omega(p(a, \theta) \mid \beta_p) u(w, a \mid \beta_u),$$

where $\beta = (\beta_p, \beta_u)$, $\omega : [0,1] \mapsto [0,1]$ is the probability weighting function parametrized by β_p, and $u : \mathbb{R}_+ \times [0,1] \mapsto \mathbb{R}$ is the utility function parametrized by β_u. In the experiment, the utility function u is $u(y, a \mid \beta_u) = v(y - c(a) \mid \beta_u)$. The cost function c and the probability of success function p are defined, as before, by equations (1.2) and (1.1), respectively.

I assume that ω takes the two-parameter Prelec form (Prelec, 1998), which is frequently used in applied work (Filiz-Ozbay, Guryan, Hyndman, Kearney, and Ozbay, 2015; Wilcox, 2015a) and has been show to have good empirical properties (Stott, 2006), with $\beta_p = (\alpha, \psi)$

$$\omega(p \mid \beta_p) = \exp(-\psi(-\ln p)^\alpha),$$

where $\psi > 0, \alpha > 0$. I also assume that v takes the standard constant relative risk aversion (CRRA) form, with $\beta_u = \gamma$

$$v(x \mid \beta_u) = \frac{x^{1-\gamma} - 1}{1-\gamma}.$$

I estimate a representative agent model on the pooled data, and use i to index individual observations. In each observation, effort is chosen to maximize $U(a \mid \delta_i, \beta)$, and the optimal effort as a function of δ and β is denoted $a^*(\delta, \beta)$. No closed-form expression for $a^*(\delta, \beta)$ exists for a given model specification, therefore, I rely on a numerical solution for optimal effort. I assume that the observed effort follows

$$a_i = a^*(\delta_i, \beta) + \varepsilon_i,$$

where ε_i is a mean-zero error term. To estimate the model, I use a non-linear least squares estimator,[23] which minimizes the sum of squared deviations between the observed and predicted choices:

$$\hat{\beta} = \arg\min_{\beta} \sum_{i=1}^{N} (a_i - a^*(\delta_i, \beta)).$$

The risk-neutral parameter vector $\beta = (0, 1, 1)$ is used as a starting value. Table 1.7 shows the estimation results.

Table 1.7: Estimates of RDU Model From the Main Task

Parameter	Estimate	SE	2.5%	97.5%
γ	0.978	0.117	0.748	1.208
α	1.603	0.071	1.464	1.743
ψ	0.821	0.054	0.716	0.926

N obs = 1625

The estimates show that, on average, subject are moderately risk averse in terms of the contribution of the curvature of the utility function to risk aversion. This finding is consistent with the previous findings in the laboratory experiments (Holt and Laury, 2002; Andersen, Harrison, Lau, and Rutström, 2008), however the estimate for the CRRA parameter is higher than is typically re-

[23] One could estimate the model using Maximum Likelihood estimator (MLE).

ported (Harrison and Rutström, 2008)[P. 121]. The 95% confidence interval for γ covers the value of one, which implies a special case of a logarithmic utility.

The estimate of α, which determines the shape of the probability weighting function, is significantly greater than one and leads to an S-shaped probability weighting. The S-shaped probability weighting implies likelihood sensitivity: subjects underweight the likelihoods of rare outcomes. As highlighted by the theoretical analysis in Section 1.3, such a shape arises precisely to fit the inverse-U pattern of effort response to difficulty that is observed in the data. Figure 1.10 (left panel) shows the probability weighting function with the parameters equal to the estimated values of α and ψ. The underweighting occurs for the probabilities roughly less than 0.25. Probabilities greater than 0.25 are overweighted. The right panel of Figure 1.10 shows the implied decision weights from the estimated probability weighting function. The decision weights are computed using *equiprobable* lotteries with different number of prizes (two, three, or four). The picture shows that, for given estimates, extreme outcomes are underweighted when there are more than two prizes, and the worst (best) outcome is underweighted (overweighted) when there are two prizes. Of course, these results reflect the equal probabilities assumed here.

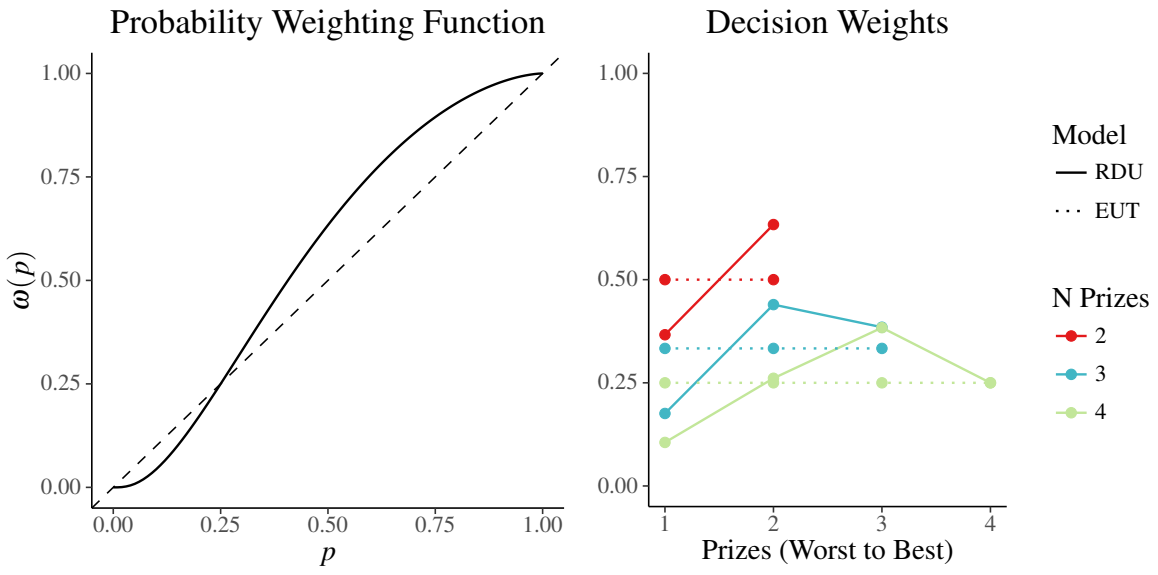


Figure 1.10: Estimated Probability Weighting Function and Implied Decision Weights from Equiprobable Lotteries (Effort Task)

To explore individual heterogeneity in preference parameters, I estimate the model for the effort task allowing for parameter heterogeneity in terms of demographics. In particular, I assume that each behavioral parameter β^j can be written as a linear combination of demographic indicators for gender and race, with Black male being the base category:

$$\beta^j = \beta^j_{Constant} + \beta^j_{Female}\mathbb{I}(Gender_i = Female) + \\ + \beta^j_{White}\mathbb{I}(Race_i = White) + \beta^j_{Asian}\mathbb{I}(Race_i = Asian).$$

Table 1.8 presents the estimation results, which reveal preference differences between demographic groups. Females tend to have a higher estimate of the coefficient of CRRA than males, while Whites tend to have a lower estimates than Blacks or Asians. There are no significant differences in the coefficient of CRRA between Blacks and Asians. Turning to the estimates of the probability weighting function, females tend to have a slightly higher estimate of the shape parameter α than males, though the difference is not statistically significant. Similarly, there are no statistically significant differences in the estimates of α between racial groups. The estimates of the scale parameter ψ show that females tend to have lower estimates than males, while Whites tend to have higher estimates than Blacks. There are no statistically significant differences in the estimate of ψ between Asians and Blacks or Asians and Whites.

Figure 1.11 interprets these estimates graphically by showing the estimated probability weighting functions and implied decision weights from equiprobable lotteries for males and females separately. For males, the probability weighting function is clearly S-shaped, while for females, it is closer to being simply concave. As a result, males tend to overweight outcomes between the extremes, while females tend to overweight best outcomes.

Figure 1.12 presents similar results for races. The differences between races are less pronounced than the differences between genders, which results in similar S-shaped probability weighting functions and similar decision weights.

Figure 1.13 shows the actual and predicted mean effort choices for the pooled data across the different values of the four treatment variables. The model fits the comparative statics found in

Table 1.8: Estimates of RDU Model from Main Task with Demographic Covariates

Parameter	Estimate	SE	2.5%	97.5%
γ				
Constant	0.528	0.150	0.234	0.821
Female	1.339	0.388	0.577	2.100
White	−1.375	0.486	−2.327	−0.422
Asian	0.250	0.193	−0.129	0.629
α				
Constant	1.479	0.084	1.314	1.644
Female	0.121	0.174	−0.219	0.461
White	−0.204	0.195	−0.586	0.178
Asian	0.034	0.148	−0.256	0.325
ψ				
Constant	1.075	0.079	0.921	1.230
Female	−0.796	0.095	−0.982	−0.611
White	0.615	0.161	0.300	0.930
Asian	0.222	0.114	−0.001	0.446

N obs = 1625.

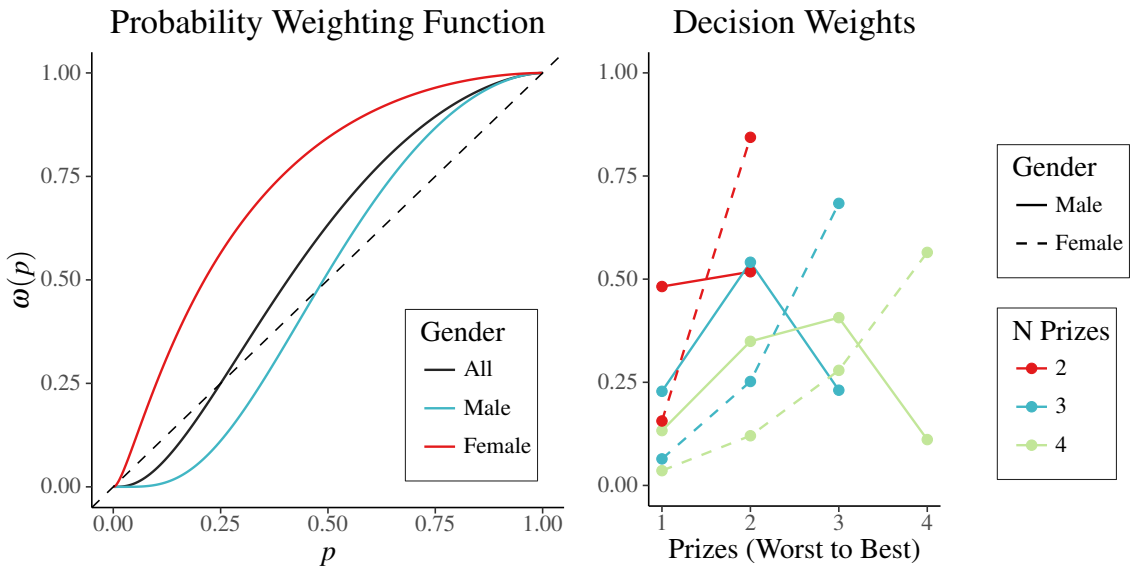


Figure 1.11: Estimated Probability Weighting Functions and Implied Decision Weights from Equiprobable Lotteries by Gender (Effort Task)

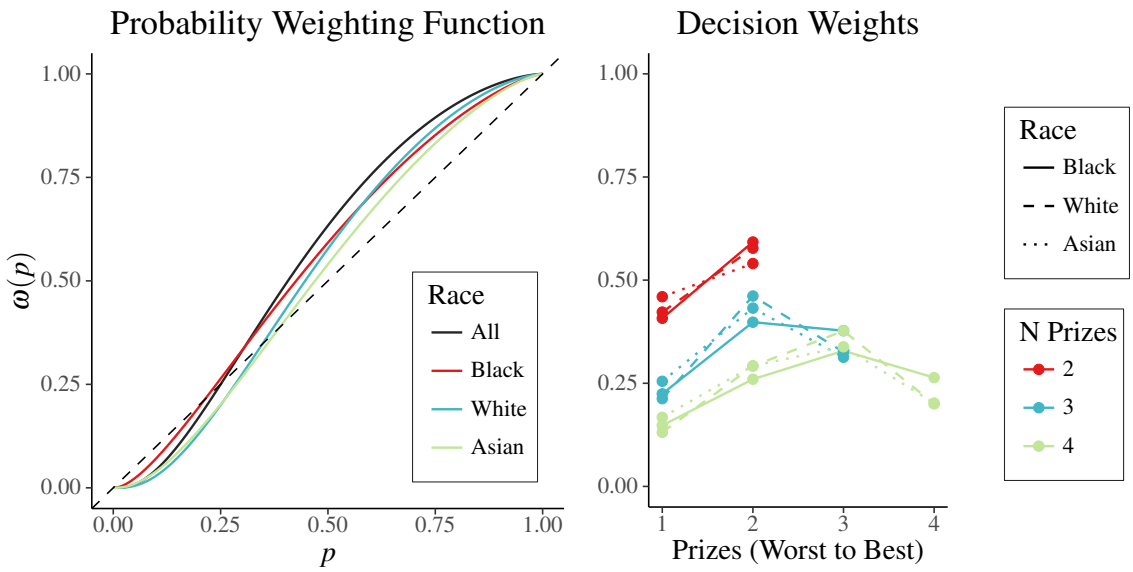


Figure 1.12: Estimated Probability Weighting Function and Implied Decision Weights from Equiprobable Lotteries by Race (Effort Task)

the data reasonably well. In particular, it does capture the inverse-U pattern of effort response to difficulty. It overpredicts, however, the mean effort for $\theta = 0.75$ leading to a more prolonged increase in effort as difficulty increases and a sharper drop in effort when difficulty changes from 0.75 to 1.

Given that the estimated shape of the probability weighting function differs from the inverse-S shape reported in some studies,[24] I ask whether this can be explained by the differences between subject pools. To explore this possibility, I estimate the same RDU model on a different risk task: binary lottery choices. If the RDU estimates from this auxiliary task turn out to be similar to the estimates from the main effort task, that would imply that the differences in estimates can in fact be attributed to the differences between subject pools. On the other hand, if the RDU estimates from the auxiliary risk task are close to what is sometimes reported in the literature, a different explanation is needed.

In the binary lottery choice task, one observes a sequence of pooled lottery choices $\{l_i\}_{i=1,\ldots,N} \in \{l_i^L, l_i^R\}$. In each pair, l_i^L is the lottery that appears on the left side of the choice screen, and l_i^R is

[24] As noted earlier, while the inverse-S weighting is a common assumption, some studies report significant deviations from it for many subjects (Wilcox, 2015b).

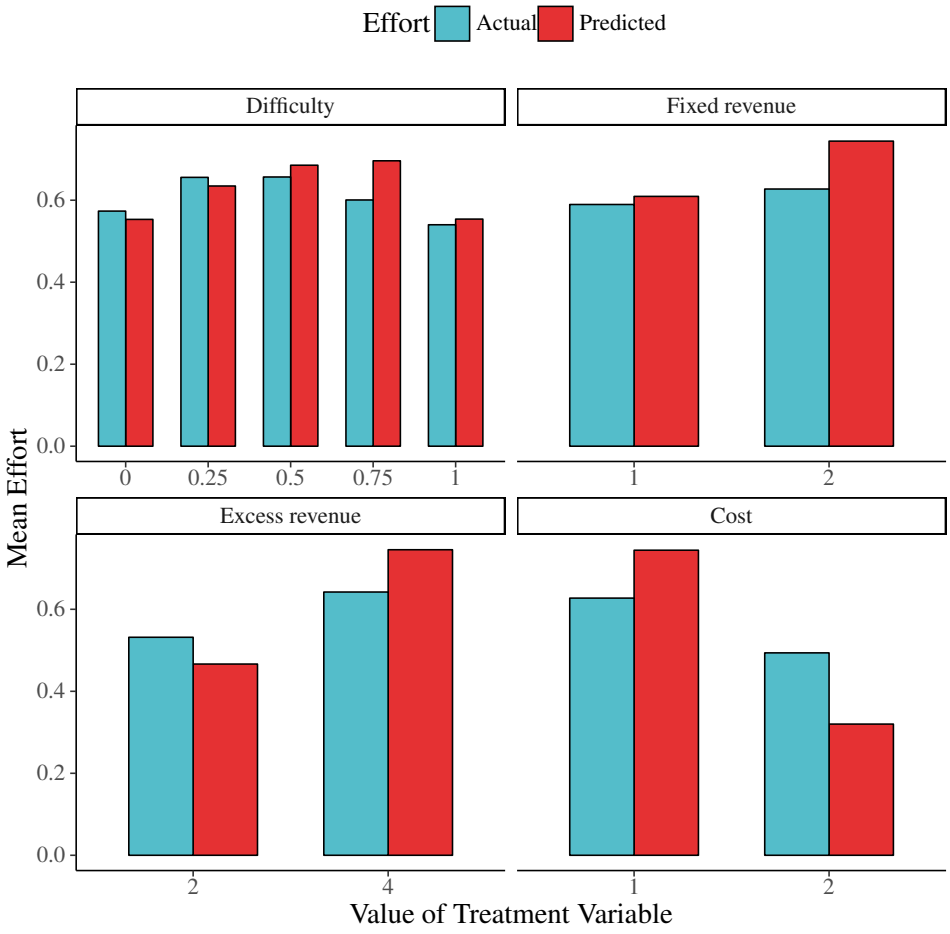


Figure 1.13: Actual and Predicted Mean Effort Levels

the lottery that appears on the right side of the screen. For a representative agent with a preference parameter β, the rank-dependent utility of a lottery l is

$$\begin{aligned} U(l \mid \beta) = & \\ & = \sum_{i=1}^{k} \left(\omega \left(q_{(1)}(l) + \ldots + q_{(i)}(l) \mid \beta_p \right) - \omega \left(q_{(1)}(l) + \ldots + q_{(i-1)}(l) \mid \beta_p \right) \right) \times \\ & \times u \left(x_{(i)}(l) \mid \beta_u \right), \end{aligned}$$

where $\beta = (\beta_u, \beta_p)$, each lottery l has k outcomes ranked from the highest, $x_{(1)}$, to the lowest, $x_{(k)}$, with the corresponding (non-zero) probabilities $q_{(1)}, \ldots, q_{(k)}$. The probability weighting function $\omega : [0,1] \mapsto [0,1]$ is parametrized by β_p, and the utility function $u : \mathbb{R}_+ \mapsto \mathbb{R}$ is parametrized by β_u. I assume, as before, that ω takes a two-parameter Prelec form (with both parameters being positive), and that u takes the constant relative risk aversion form.

I use the contextual utility (Wilcox, 2011) to model the likelihood of choosing a particular lottery. The benefit of this stochastic model, as opposed to the standard strong utility (logit) model, is that it allows for heterogeneity in the sensitivity to the utility differences across the lottery pairs and preserves the notion of being more risk averse in a stochastic setting. It also has been shown to provide a good out-of-sample prediction power (Wilcox, 2015a). The likelihood of choosing l^R under this model is

$$p(l^R \mid \beta, \mu) = \Lambda \left(\frac{U(l^R \mid \beta_u, \beta_p) - U(l^L \mid \beta_u, \beta_p)}{\mu \left(u(x^h \mid \beta_u) - u(x^l \mid \beta_u) \right)} \right),$$

where x^h and x^l are the highest and the lowest payoffs in the lottery pair, and $\mu > 0$ is the noise parameter that determines how sensitive the choice likelihoods are to the utility differences (low μ meaning very sensitive, and high μ meaning not sensitive). The likelihood of choosing l^L is $1 - p(l^R \mid \beta, \mu)$.

I estimate the model by maximizing the log-likelihood function

$$(\hat{\beta},\hat{\mu}) = \arg\max_{\beta,\mu} \sum_{i=1}^{N} \Big(\mathbb{I}(l_i = l^R) \ln p(l_i^R \mid \beta,\mu) + \mathbb{I}(l_i = l^L) \ln p(l_i^L \mid \beta,\mu) \Big),$$

where $\beta = (\beta_u, \beta_p)$, $\beta_u = \gamma$, $\beta_p = (\alpha, \psi)$, and $\mathbb{I}(\cdot)$ is the indicator function. As before, the vector of risk-neutral parameters $\beta = (0,1,1)$ is used as a starting value. Table 1.9 shows the estimation results.

Table 1.9: Estimates of RDU Model From Lottery Task

Parameter	Estimate	SE	2.5%	97.5%
γ	0.678	0.015	0.646	0.712
α	0.803	0.022	0.761	0.852
ψ	0.706	0.022	0.667	0.756
μ	0.103	0.004	0.095	0.112

Log-likelihood: -4095.24, N obs = 6664.

The estimate of γ is much lower than the corresponding estimate from the effort task, indicating that in the lottery task, the subjects are less risk averse in terms of the curvature of the utility function. This estimate, however, is closer to the estimates reported by previous studies. More importantly, the estimate of probability weighting shape α is significantly below one. This implies that in the lottery task, the subjects are characterized by an inverse-S-shaped probability weighting. Figure 1.14 (left panel) shows the probability weighting function with the parameters equal to the estimated values of α and ψ from the lottery task. The function features a heavy overweighting of probabilities less than 0.8, while probabilities greater than 0.8 are slightly underweighted. The right panel of Figure 1.14 shows the implied decision weights from the estimated probability weighting function in the lottery task. The decision weights, as before, are computed using equiprobable lotteries with different number of prizes (two, three, or four). The picture shows that extreme outcomes are overweighted when there are more than two prizes, and that the worst (best) outcome is underweighted (overweighted) when there are two prizes. As before, these results reflect the equal probabilities assumed here.

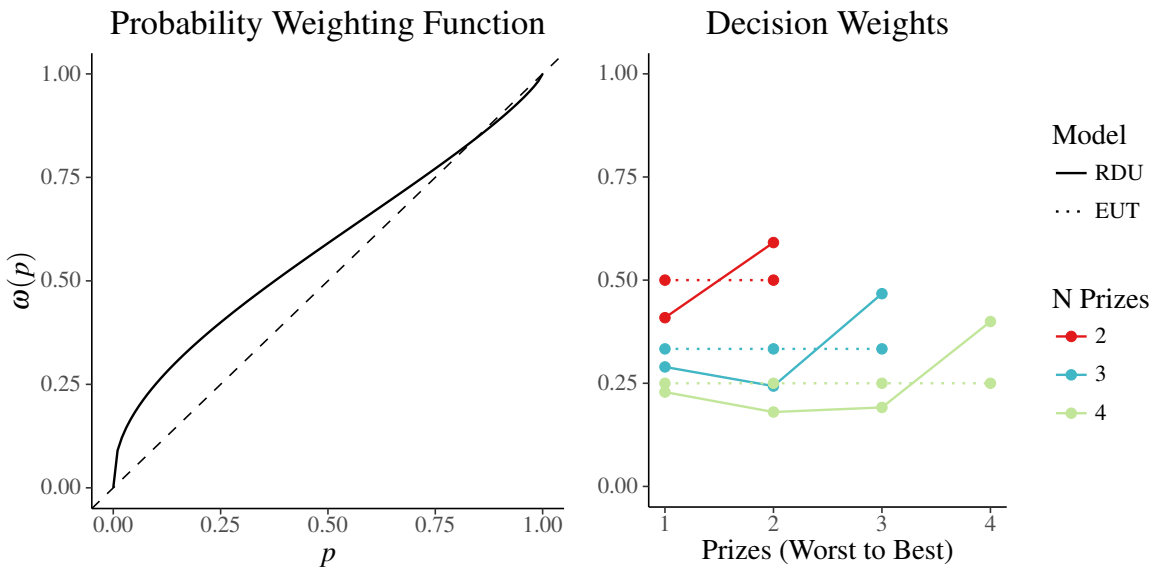


Figure 1.14: Estimated Probability Weighting Function and Implied Decision Weights from Equiprobable Lotteries (Lottery Task)

The question is then what features of the effort task make it different from the lottery task so that the estimated shapes of the probability weighting vary so dramatically between them. The first obvious difference is that the effort task is framed in a specific context (i.e., work on a project, task difficulty) while the lottery task is framed in an abstract context. Contextual instructions can affect subjects' behavior, as is evident from a variety of studies (Harrison and List, 2004; Alekseev, Charness, and Gneezy, 2017).[25]

Another difference between the two tasks is the frequency of feedback. In the effort task, the feedback on the choice outcome was provided every round, while in the lottery task, the feedback was provided only after all the choices were made. Evidence from incentivized laboratory experiments suggests that providing feedback can make subjects more sensitive to probabilities resulting in linear probability weighting (Van de Kuilen, 2009), or even in S-shaped probability weighting (Hertwig, Barron, Weber, and Erev, 2004; Hertwig and Erev, 2009) in a binary lottery choice setting, which is consistent with the present findings.[26]

[25]It is worth emphasizing that even though the use of context in the effort task might have affected subjects' behavior, this should not be viewed as a shortcoming of the design. The labor context is precisely the context, behavior in which is of interest.

[26]The math quiz that was distributed before the choice tasks could have potentially stimulated the subjects to think more carefully about probabilities. However, this effect would apply to both tasks and not just the effort task.

These considerations suggest that further treatments should manipulate the context of the effort task and the provision of feedback. The two tasks differ in other aspects such as the layout of choice screens and the number of alternatives to choose from. These difference, however, are unlikely to generate such dramatic differences in estimated preference parameters.

1.5 Conclusion

Recent research on the incentive effects and workers' effort highlights the role of alternative behavioral tools and suggests studying more comprehensive incentive schemes that go beyond mere monetary rewards. The present work contributes to this trend by studying the effect of task difficulty on workers' effort and comparing it to the effect of monetary rewards.

To do so, I set up a tightly controlled laboratory experiment, in which subjects choose their effort levels in projects with a binary stochastic outcome. Higher effort results in a higher probability of success and higher revenue, which is a sum of conditional and unconditional reward, but leads to a higher monetary cost of effort that is subtracted from the revenue. A project's difficulty is modeled as a variable that enters the probability of success function, along with effort, and that has a negative impact on the success probability.

I find that difficulty has an inverse-U effect on effort: effort first increases as difficulty goes up, reaches a peak, and then drops as difficulty continues to increase. The effects of monetary rewards are in line with the previous findings: higher conditional rewards and lower cost lead to higher effort, while higher unconditional rewards have only a weakly positive effect on effort. The effect of difficulty on effort is quantitatively large, the difference between the highest and lowest effort values across the entire range of difficulty is comparable to the effect of doubling conditional rewards. Difficulty acts as a mediator of monetary rewards: they are most effective when difficulty is set to a medium level. I uncover significant heterogeneity in the magnitudes and types of effort responses to difficulty and monetary rewards. Males are more likely to exhibit the inverse-U pattern of effort response to difficulty than females.

I theoretically show that the inverse-U effect of difficulty is inconsistent with the EUT model. It can be generated, however, by the RDU model that allows for non-linear probability weighting. I estimate a structural RDU model and find the evidence for an S-shaped probability weighting function that implies that subjects underweight rare outcomes, while an inverse-S specification assumes that subjects overweight rare outcomes. I re-estimate the model on the data from the same subjects but in an alternative binary lottery task and find evidence for inverse-S-shaped probability weighting. I discuss the possible causes of the differences in the estimates, among which are the use of work-related context in the effort task and the frequency of feedback provision.

These findings suggest that task difficulty is an important and costless behavioral tool that should be taken into account when designing the optimal incentive schemes for workers. In particular, setting task difficulty at a medium level[27] so that the task is reasonably challenging can boost effort exertion and amplify the positive effects of conditional and unconditional rewards.

[27] While this is true for most subjects in this experiment, one still has to be mindful about possible heterogeneity in responses, especially between males and females.

Chapter 2

Success Decomposition: Using Response Times to Measure Ability and Motivation

2.1 Introduction

What determines success in life: cognitive abilities, character skills, or some combination of both? This is an important question in the economics of human development, the answer to which can shape the funding for government policies aimed at encouraging one type of characteristics versus the other. The literature in this area, both observational and experimental, initially focused on studying the relationship between cognitive abilities and life outcomes and found that cognitive ability is an important determinant of earnings (Murnane, Willett, and Levy, 1995), risk and time preferences (Dohmen, Falk, Huffman, and Sunde, 2010), and the quality of decision-making (Agarwal and Mazumder, 2013). Increasingly, however, researchers investigate the role of character skills. This literature shows that character skills play an important, and often dominant, role in determining various outcomes, such as life-time earnings (Heckman, Stixrud, and Urzua, 2006; Heckman, Pinto, and Savelyev, 2013), teenage pregnancy, marital status, smoking, and engaging in criminal activities (Duckworth, Quinn, Lynam, Loeber, and Stouthamer-Loeber, 2011), and strategic reasoning (Gill and Prowse, 2016).

The answer to the question of the relative importance of cognitive abilities versus character skills depends on their valid measurement. A typical approach in the literature uses performance on a cognitive test as a measure of ability, and self-reported statements to a questionnaire (e.g., Big Five) as measures of character skills (Heckman, Stixrud, and Urzua, 2006). Two problems arise with this approach. First, test performance never reflects cognitive ability alone, it also reflects motivation. The standard approach thus confounds ability with the combination of ability, motivation, and potentially other character skills, resulting in incorrect conclusions about the role of ability. Using test performance as a measure of ability is unwarranted on theoretical and empirical grounds (Duckworth, Quinn, Lynam, Loeber, and Stouthamer-Loeber, 2011; Segal, 2012).[1] Using test performance as a noisy proxy for ability could potentially be acceptable if the "noise" content of test performance relative to the "signal" (i.e., ability) content were small. The existing literature, however, has little to say about the magnitude of the "signal" component in test performance. Second, using self-reported statements for gauging character skills is problematic, since it is hard to assess their validity. They could represent wishful thinking rather than true characteristics (Borghans, Duckworth, Heckman, and Ter Weel, 2008). Additionally, self-reports about desirable social characteristics, such as intellect or motivation, may be intentionally distorted because of the self-image concerns (Ewers and Zimmermann, 2015).

I propose a novel method of measuring cognitive ability and character skills (motivation). My method improves upon the existing approaches by *a*) decomposing the effects of ability and motivation on test scores and *b*) providing an objective measure of motivation. My method is based on a dynamic stochastic model of optimal effort choice with ability and motivation being the structural parameters of the model. I show how these parameters can be estimated from the data on outcomes and response times in a cognitive task using a version of a threshold regression model (Lee and Whitmore, 2006).

[1] For example, consider two students, Adam and Bob, who are taking a cognitive test. Adam has high cognitive ability but is not interested in the outcome of the test. Bob, on the other hand, has lower cognitive ability but is highly motivated to get the right answers. As a result, Bob might end up having a higher score on the test, which according to the standard approach would imply that Bob has higher ability than Adam, while in reality their ranking by ability is the opposite. Appendix B.1 formalizes this intuition.

My approach is based on an explicit procedural modeling of the decision-making process and is inspired by the literature in cognitive psychology on sequential-sampling, or drift-diffusion, models (Ratcliff, 1978; Busemeyer and Townsend, 1993). These models have been shown to perform well in jointly predicting outcomes and response times (Gold and Shadlen, 2007). I use response times (i.e., the time needed to give an answer to a question, such as an item in a cognitive test) as a proxy for effort, following Wilcox (1993) and Ofek, Yildiz, and Haruvy (2007). An agent's effective effort is modeled as a Brownian motion with drift, where the drift represents the agent's *ability*. Higher ability leads to a more rapid effort accumulation. The accumulated effective effort at any point in time determines the probability to answer a question successfully. Thus agents with high ability would reach higher probability of success (or high performance score, over many trials) in a given amount of time than agents with low ability. Success on a given trial yields a utility to the agent that represents her *motivation* (intrinsic, or extrinsic, or both). Effort is costly, and the more time an agent spends on a task, the higher will be the accumulated cost of effort. The agent's problem is, therefore, to choose the optimal moment to stop the effort accumulation process that balances the cost of additional effort versus the potential gain of further accumulation. The solution to the agent's problem takes the form of a threshold rule in terms of the accumulated effective effort. I provide a closed-form solution for the optimal threshold under certain parametric assumptions and show how it is related to the underlying parameters of the model, including ability and motivation. The biggest advantage of the proposed model is that it allows to decompose the effects of ability and motivation on test scores and to provide an objective measure of motivation. The proposed model is also an example of how explicit procedural modeling of response times, as advocated by Spiliopoulos and Ortmann (2018), can be used constructively to solve problems in economic settings.

Another benefit of the proposed model is that it allows for a seamless transition to an econometric estimation. I show how the underlying parameters of the model can be estimated using maximum likelihood methods from the data on outcomes and response times in a series of trials of a cognitive task. I also argue that the identification of these parameters from the observed data is

transparent. The proposed estimation strategy can be viewed as a version of a threshold regression model used in survival analysis (Lee and Whitmore, 2006).

To assess the empirical validity of the proposed method and to compare its performance to the traditional approach, I conduct a laboratory experiment. In the experiment subjects go through a series of trials of a Digit-Symbol test in which they would have to match symbols to digits. This test finds widespread use in the literature (e.g., Segal (2012); Dohmen, Falk, Huffman, and Sunde (2010)) and in intelligence scales such as WAIS (Weiss, Saklofske, Coalson, and Raiford, 2010). Subjects are paid a flat amount for completion of the task. I observe the outcomes of trials (success or failure) and response times, which together with a structural model allows me to uncover subjects' ability and motivation parameters. I measure subjects' performance on a Cognitive Reflection Test Frederick (2005) and Big Five personality scores, which are popular measures of cognitive ability and character skills. I also elicit subjects' risk preferences and decision-making quality using an incentivized task that follows the design of Choi, Kariv, Müller, and Silverman (2014).

I uncover substantial heterogeneity among subjects in terms of their ability and motivation. I find that a test score is a very noisy measure of true ability and that the observed variation in test scores is mostly due to variation in motivation rather than variation in ability. Using the proposed theoretical model, I am able to simulate the counterfactual distribution of the test scores implied by the variation in one parameter only (ability or motivation). The counterfactual distribution of test scores implied by ability alone is much tighter than the actual distribution, the standard deviation of the counterfactual distribution is more than twice as low. On the other hand, the counterfactual distribution of test scores implied by motivation alone is very close to the actual distribution.

The observed variation in ability and motivation in the sample can be partially explained by the variation in subjects' demographic characteristics, measures of decision-making ability, and preferences. I find that females have higher ability than males, and that White and Asian subjects have higher ability than Black subjects. The gender effect is robust to including various additional controls, however, the race effect becomes weaker after controlling for subjects' decision-making

ability. In particular, subjects' consistency with utility-maximization in the risk task, as measured by the CCEI index[2] (Choi, Kariv, Müller, and Silverman, 2014), is strongly positively associated with higher ability. The effect of risk aversion on ability or motivation is not statistically significant. Motivation turns out to be a more idiosyncratic characteristic than ability, as I do not find strong associations between motivation and gender, race, and other control variables.

Surprisingly, I find virtually no association between estimated measures of ability and motivation and their self-reported counterparts from the Big Five questionnaire. The effects of the self-reported measures are close to zero and are very imprecisely estimated. I argue that self-reported measures may be biased to the point that they distort the true relation between objective and self-reported measures. It is well-known that people tend to be overconfident about their abilities relative to others, for example, due to self-image concerns. Similarly, subjects are found to misreport information when it is beneficial to do so in the experiments on lying (e.g, Gneezy, Kajackaite, and Sobel (2018)). It is plausible, therefore, that when asked about a socially desirable skill, such as intellect or motivation, an individual has an incentive to artificially boost his or her report about that skill. If this effect is stronger for subjects whose true characteristic is low, the self-reported measure will completely distort the positive association between the report and the truth. This suggests that the answers to the survey questions about desirable character skills may be unreliable.

This result, while suggestive of the potential issues with relying on self-reported measures of cognitive ability or character skills, is far from being conclusive. More research is required on the relationship between self-reported and objective measures of ability and motivation. One line of research should consider alternative cognitive tasks to evaluate the sensitivity of this relationship to various contexts. Another important direction of further research is to use the objective and self-reported measures as explanatory variables for relevant outcomes, such as life-time income, risky behavior, or the quality of financial decisions, similarly to what has been done by, e.g., Noussair,

[2]Appendix B.3 analyzes in more detail the estimates of CCEI in the sample, as well as the test power computed using a method by Bronars (1987). In general, the budget lines used in the risk task provide a strict test of consistency with GARP.

Trautmann, and van de Kuilen (2014) or Choi, Kariv, Müller, and Silverman (2014) in the context of risk attitudes.

Turning to the relative importance of ability versus motivation on the success on a cognitive task, I find that motivation plays a slightly bigger role than ability. In particular, subjects with high motivation and low ability tend to have higher scores than subjects with low motivation and high ability. This pattern is driven by the fact that highly motivated subjects often reach near-perfect scores, while high-ability (but not motivated) subjects never reach such scores. This result contributes to the ongoing discussion about the role and relative importance of cognitive ability and character skills in life. Existing studies show that the importance of different skills varies by context and no single skill dominates others across the board. Different character skills, furthermore, are found to be more or less important depending on context. Barrick and Mount (1991) conduct a meta-analysis of the effect of the Big Five character skills on job performance in different occupations. They find that conscientiousness is a significant predictor of all job performance measures across all occupations. The importance of the other four character skills, however, varies by measures and occupation. On the other hand, Choi, Kariv, Müller, and Silverman (2014) do not find a significant effect of conscientiousness on subjects' wealth[3] in a field experiment in Netherlands. Chetty, Friedman, Hilger, Saez, Schanzenbach, and Yagan (2011) analyze data from Project STAR and find evidence that improvement in character skills early in life has a long-lasting impact on future earnings. Similarly, Heckman, Pinto, and Savelyev (2013) use the data from the Perry Preschool program and find that induced changes in character skills explain a larger proportion of adult outcomes than induced changes in cognitive ability. Nilsson (2017) exploits an exogenous variation in alcohol availability for parents in Sweden to study its effect children's cognitive ability and character skills and resulting labor market and educational outcomes. This study finds that character skills play a bigger role in explaining labor market outcomes, while cognitive ability plays a bigger role in explaining educational outcomes. Gill and Prowse (2016) compare the effects of cognitive ability and character skills on the quality of strategic decision-making in a beauty

[3]The study defined wealth as a sum of net worth of all household members, which included checking and savings accounts, stocks and bonds, real estate, business assets, and loans and mortgages.

contest game. They find that cognitive ability has a higher effect than the character skill that combines agreeableness and emotional stability. Other character skills, such as conscientiousness or openness, are found to have no effect on the quality of strategic decision-making.

The proposed method complements alternative approaches designed to infer ability and motivation from observable data. The first approach, due to Heckman *et al.* (2006), is based on the idea that cognitive ability and personality traits can be measured using separate tests designed to disentangle them. Then a factor model can be used to infer the underlying ability factor from a series of cognitive tests and the underlying personality traits factor from a series of personality tests.[4] The downside of this approach, however, is the assumption that cognitive test scores depend only on ability and not motivation, which does not hold in practice.[5] Any performance score reflects both qualities, though in different degrees. In addition to that, personality traits measured by personality tests, such as Big Five, are self-reported and there is no objective way of knowing whether subjects manipulate the test or not (Borghans, Golsteyn, Heckman, and Humphries, 2011). The second approach relies on exogenous variations in experimental conditions, such as incentives, time limits and task difficulty. Segal (2012) varies the level of incentives offered in a cognitive task and uses the analysis of variation to disentangle the impact of ability and motivation on the test scores. Borghans, Meijers, and Ter Weel (2013) vary the levels of incentives, time limits allowed to answer the questions, and task difficulty. They build a structural model of effort exertion and use their exogenous variations to infer the ability and both intrinsic and extrinsic motivation of subjects. McDaniel and Rutström (2001) vary a penalty for excess moves in a Tower-of-Hanoi game and find that higher penalty results in more time spent on a task but does not affect performance.

[4] In a factor model, a researcher observes a series of outcomes that are assumed to be associated with one or more unobserved covariates. The model then estimates the unobserved factors, which are typically fewer than the number of observed outcomes. In applications to cognitive ability, for instance, the ability factor is estimated from a series of test scores.

[5] See Borghans, Duckworth, Heckman, and Ter Weel (2008) for a summary of empirical studies that show significant positive effects of incentives on performance on cognitive tests. Also, see Appendix B.1 for a theoretical argument.

2.2 Theoretical Framework

Consider an agent working on a task, such as a question on a cognitive test. The outcome of the task is binary, success or failure, and the agent accumulates costly effort over time to improve her chances of success. Given an initial level $\tilde{E}_0$,[6] the accumulated effective effort will evolve according to the following stochastic process:[7]

$$d\tilde{E}_t = \alpha \varepsilon_t dt + \sigma dW_t, \tag{2.1}$$

where dW_t is the increment of the Wiener process with a standard deviation σ, ε_t is an instantaneous effort intensity, and α is the agent's ability parameter. Given the assumption about the effort accumulation process, higher ability will result in a more rapid effort accumulation, so that in a given period of time an agent with higher ability will accumulate more effective effort and reach a higher probability of success than an agent with lower ability.[8] The accumulated effective effort is transformed to a probability of success via a function $p : \mathbb{R}_+ \mapsto [0,1]$ assumed to be increasing and concave. The concavity of p implies that each additional unit of effort (time) spent produces less output in the form of a probability of success. Specifically, I assume that the probability of success function takes the form of an exponential CDF: $p(\tilde{E}_t) = 1 - e^{-\tilde{E}_t}$.[9]

The success on a task yields the agent utility μ, which represents her motivation to work on the task. Motivation can be comprised of both intrinsic (the natural desire to succeed or a pleasures derived from completing the task correctly) and extrinsic (such as a piece-rate) parts. The agent

[6] The initial level $\tilde{E}_0$ captures the case when a task has multiple choices to choose from, so that there is a non-zero probability of getting the task correctly simply by guessing.

[7] A natural requirement for the accumulated effective effort is that it should be non-negative, $\tilde{E}_t \geqslant 0, \forall t$. This requirement can, in principle, be violated for the Brownian motion with drift. However, under a certain reasonable range of parameter values the probability of violating this requirement will be very small. An alternative would be to the Geometric Brownian motion with drift, but in this case one would have to specify a non-negative starting point for the process. It is not obvious what this starting point should be in an open-ended task, while under the Brownian motion with drift the starting point can be conveniently set to zero.

[8] Note that this does not mean that subjects with higher ability are less deliberate. This statement only means that for an exogenously set amount of time an agent with higher ability will accumulate more effort. However, since response time is a choice variable, it will be affected by ability in the optimum.

[9] This can be thought of as a cognitive production function with diminishing marginal returns and no fixed costs. The no-fixed-costs assumption makes sense in the context of the present experiment, in which the task at hand does not require prior knowledge, but it may not be true in other contexts.

then solves the following dynamic stochastic optimization problem, which takes the form of an optimal stopping problem:

$$\max_{\tau} \mathbb{E}\left[\int_0^T -c(\varepsilon_t)e^{-\rho t}dt + \mu p(\tilde{E}_\tau)e^{-\rho\tau}\right], \tag{2.2}$$

subject to the constraint given by (2.1). The agent's aggregate utility function has two components. The first component is the discounted expected utility, $\mu p(\tilde{E}_\tau)e^{-\rho\tau}$, of obtaining value (motivation) μ with probability of success $p(\tilde{E}_\tau)$ that depends on the accumulated effort at time t.[10] The second component is the accumulated discounted cost of effort that results from exerting instantaneous effort ε_t over time.

At each point in time the agent's decision can be thought of as the decision to either stop the effort accumulation process or continue to accumulate effort. Suppose the agent finds herself at a point where the accumulated effective effort is $\tilde{E}_t$. She has to consider the benefit of the immediate termination of the effort process, $\Omega(\tilde{E}_t) = \mu p(\tilde{E}_t)e^{-\rho t}$, against the additional expected utility she would get by following the optimal decision rule. This additional utility comes from the fact the the stochastic process can go up, given a non-negative drift, and thus increase the probability of success.

The continuation of the effort accumulation incurs the costs of the additional exerted effort, $c(\varepsilon_{t+\Delta t}) - c(\varepsilon_t)$. I assume that these additional costs are time-invariant, i.e., depend only on the time shift Δt and not on the time point t, which, together with exponential discounting and additive separability,[11] simplifies the decision problem by making the decision rule independent of time.[12] Therefore, I use the assumption of a linear functional form for the cost function, $c(\varepsilon) = k\varepsilon$. Since

[10] Exponential discounting is used as a way to preserve the tractability of the model. This assumption makes the decision problem stationary and allows for a straightforward derivation of a closed-form solution. For the same reason, it is assumed that the agent maximizes expected utility. It would be an interesting extension of the model to consider the possibility of a probability weighting that would allow subjects to have heterogeneous beliefs about their probabilities of success.

[11] Additive separability is a convenient assumption that makes the model more tractable. This assumption, however, might be too strong in some settings (Andersen, Harrison, Lau, and Rutström, 2018).

[12] I also implicitly assume here that the agent does not have a time constraint T. This would lead to a finite-horizon problem, which, in general, does not have an analytical solution.

the effort intensity is unobserved and cannot be estimated from the data on outcomes and response time, I make a simplifying assumption that $\varepsilon_t \equiv 1$.

In this setting the solution to the agent's problem (2.2) is a stopping rule of the form:

$$\begin{cases} \text{stop}, & \tilde{E}_t \geqslant \tilde{E}^*, \\ \text{continue}, & \tilde{E}_t < \tilde{E}^*, \end{cases}$$

where $\tilde{E}^*$ is the optimal threshold level of accumulated effective effort that depends on the parameters of the problem. This optimal threshold can be found using the Hamiltonian-Jacobi-Bellman (HJB) equation of the problem:

$$0 = -\rho h - k + \alpha h' + \frac{\sigma^2}{2} h'', \tag{2.3}$$

where h is the value function of the problem.

Using the HJB equation (2.3) along with the appropriate boundary conditions leads to the following result:

Proposition 4. *The optimal threshold is given by*

$$\tilde{E}^* = \ln\left(1 + \frac{1}{\beta}\right) - \ln\left(1 + \frac{k}{\mu\rho}\right), \text{ where } \beta \equiv \frac{-\alpha + \sqrt{\alpha^2 + 2\rho\sigma^2}}{\sigma^2}. \tag{2.4}$$

2.3 Estimation Strategy

As a first step, I reduce the dimension of the underlying parameters of the model, $(\alpha, \mu, \sigma, k, \rho)$ by making the normalization $k = 1$ and the assumption $\rho = 0.01$ for all subjects. The normalization comes from the fact that the optimal threshold depends only on the ratio of μ to k, and therefore the two parameters cannot be estimated separately. The assumption about the discounting factor is made to simplify the inference and since the available data is unlikely to be able one to uncover

meaningful differences in time preferences among subjects over the small intervals of time[13] during which the experiment takes place.

In order to estimate the remaining parameters of interest, (α, μ, σ), the data on both the inputs and outputs of the decision process are required.[14] Suppose that a sequence of N trials of a decision task performed by a subject is observed. Each observation consists of a pair (x_i, t_i), $i = 1, \ldots, N$, where x_i is a realization of a Bernoulli random variable, which indicates whether the task was solved correctly of not, and t_i is the time it took to reach a decision.

The joint likelihood of observing an outcome x_i after time t_i has passed is given by

$$l(x_i, t_i \mid \alpha, \mu, \sigma, k) = p(\tilde{E}^*)^{x_i}(1 - p(\tilde{E}^*))^{1-x_i} f(t_i).$$

This likelihood consists of the two parts, one of which is the Bernoulli likelihood, and the other is the likelihood of a stochastic process hitting a threshold. In this formula, the probability of success p is a function of the optimal threshold $\tilde{E}^*$, which, in turn, depends on the underlying parameters $(\alpha, \mu, \sigma, k, \rho)$ through the equation (2.4).

The likelihood $f(t_i)$ comes from the properties of the stochastic process. Let $T_{\tilde{E}^*} = \inf\{t : \tilde{E}_t = \tilde{E}^*\}$ be the time when the accumulated effective effort given by (2.1) first hits the optimal threshold $\tilde{E}^*$. This time is a random variable with the pdf given by

$$f(t \mid \tilde{E}^*, \tilde{E}_0, \alpha, \sigma) = \frac{\tilde{E}^* - \tilde{E}_0}{\sqrt{2\pi\sigma^2 t^3}} \exp\left(-\frac{(\tilde{E}^* - \tilde{E}_0 - \alpha t)^2}{2\sigma^2 t}\right).$$

This expression depends on the underlying parameters both directly and indirectly through the optimal threshold $\tilde{E}^*$.

The model can be estimated using the maximum likelihood method:

$$(\hat{\alpha}, \hat{\mu}, \hat{\sigma}) = \arg\max_{\alpha,\mu,\sigma} \ln \mathscr{L}(\alpha, \mu, \sigma \mid \mathbf{x}, \mathbf{t}) \equiv \sum_{i=1}^{N} \ln l(x_i, t_i \mid \alpha, \mu, \sigma). \tag{2.5}$$

[13] Provided that discounting is exponential. This might not be true for quasi-hyperbolic discounting.

[14] In the application below, I use the transformation $\tau \equiv 1/\sigma$ to obtain a measure I call consistency. Higher consistency implies a lower variation in response times.

Note that the original problem involved effort intensity ε_t. In order to avoid confounding higher effort intensity with higher ability, I assume that there is no subject heterogeneity with respect to the effort intensity.[15] This assumption implies that one can normalize the effort intensity parameter to 1 for every subject.

The econometric model can be viewed as an extension of a Threshold Regression model (Lee and Whitmore, 2006). This model describes time-to-event data by assuming that there is an underlying stochastic process, and the event is observed when the process hits a zero threshold. This type of models is widely used in survival analysis, and in particular, in medical data. A nice feature of this model is that the distribution of the first-hitting-times follows from the properties of the stochastic process and is well-known for some common processes.

In a typical Threshold Regression model, the stochastic process is a Brownian motion with a negative drift that starts at some positive value. The diffusion parameter of the process is usually normalized to one, since the units of the process are defined only up to a scale. This normalization allows identifying the other two parameters: the starting value of the process and the drift parameter. In the present application, this procedure is slightly modified to accommodate the features of the theoretical model. The underlying stochastic process is a Brownian motion with a positive drift, which starts at $\tilde{E}_0$. The decision is made and its outcome is observed when the process hits the threshold $\tilde{E}^*$. In addition to the data on the time-to-event, we also observe the outcome of an event, i.e., whether the trial was a success or a failure.

Even though I estimate the parameters jointly, it is insightful to think of the estimation procedure as a sequence of small steps. These steps show clearly where the identification of each parameter comes from. On the first step, the data on the outcomes of trials allow inferring what was the probability of success. This probability remains constant in every trial, since it depends only on the underlying parameters that are assumed to remain constant across the trials. On the

[15] Focusing on the extensive margin of effort (response time) is typical for economic studies (Wilcox, 1993; Ofek, Yildiz, and Haruvy, 2007). In psychology, however, researchers come up with elaborate ways to measure effort intensity (Richter, Friedrich, and Gendolla, 2008). In the present setting, focusing on response times rather than on effort intensity is justified by the nature of the design. Allowing subjects to work on the cognitive task with no time constraints encourages subjects to use extensive margin of effort. On the other hand, if the task had a binding time constraint, subjects would be more likely to rely on the intensive margin of effort.

second step, the functional form of the probability of success pins down what was the optimal threshold, which is also constant throughout the trials. On the third step, the data on decision times combined with the inferred optimal threshold yields the estimates of ability and consistency. On the last step, these estimates, together with the estimate of the optimal threshold and the functional form given by (2.4), pin down the remaining unknown parameter, motivation.

2.4 Experiment

2.4.1 Procedures

The experiment was conducted at the Experimental Economics Center (ExCEN) lab at Georgia State University (GSU) in June 2017 and March-April 2018. A total of 11 session were conducted, and 192 subjects participated in the experiment. The subjects in the study were undergraduate students at GSU, who were invited to participate via email. Each subject participates only in one session. The subjects were allowed to proceed at their own pace and leave as soon as they complete the experimental tasks. Upon completion of the experimental tasks, each subject was paid privately in cash. The average earnings were $36.35 (the minimum earnings were $25, and the maximum earnings were $49.70), which included the earnings from the two incentivized tasks and a $5 show-up fee.

Table 2.1 shows the demographic characteristics of the sample. The experiment attracted almost twice a many female subjects as male subjects. The average age was around 20.6 years. The majority (64%) of subjects were Black (African American), followed by Asian (17%) and White (13%) subjects. The subjects were pretty evenly spread across years in school and majors. Students majoring in Economics, however, were underrepresented.

Table 2.1: Demographic Characteristics of the Sample

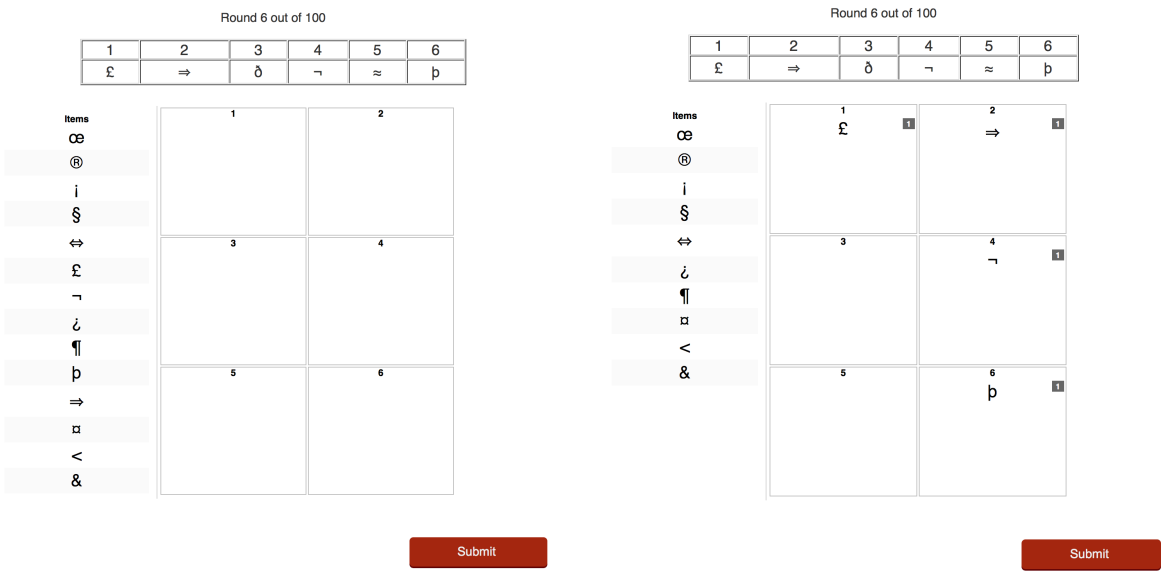
Characteristic	Mean
Gender and Age	
Male	0.33
Female	0.67
Age	20.60
Race	
Black	0.64
White	0.13
Asian	0.17
Other	0.07
Year in School	
Freshman	0.22
Sophomore	0.24
Junior	0.26
Senior	0.24
Masters	0.04
Major	
Arts	0.21
Business	0.26
Economics	0.06
STEM	0.24
Other	0.23

2.4.2 Design

Each session consisted of three parts presented to subjects in the following order: a cognitive task, a risk task, and a questionnaire. The first two tasks were incentivized, while the last one was presented in a survey format. The questionnaire included demographic questions, unincentivized questions about time preferences (as proposed by Falk, Becker, Dohmen, Huffman, and Sunde (2016)), 3-question Cognitive Reflection Test (Frederick, 2005), and personality questions.

Cognitive task

The cognitive task is a version of a Digit-Symbol Test (DST) and is based on finding correct correspondences, it consists of 100 trials.[16] In each round, subjects see six pairs of numberâĂŞsymbol combinations (the key) arranged in a table at the upper part of the screen. Below the key, there are six empty numbered boxes. See Figure 2.1a for an example.



(a) Example of a DST Round.

(b) Correctly Solved DST Round

Figure 2.1: Examples of Subject Screens in DST.

[16] Since I am estimating the underlying parameters using the maximum likelihood at the individual level, I need a large number of observations per subject. On the other hand, increasing the number of trials further would likely make subjects bored, which could affect their motivation and bias the estimates.

Subjects have to use the key to fill in the boxes with the 14 symbols located in a column to the left of the boxes. Subjects do this by dragging the symbols into the boxes. Subjects can drag a symbol back to the column or to another box, e.g., if they think that they have made a mistake. The correct solution to a given trial is straightforward. If a symbol from the column is in the key, subjects need to drag it to the corresponding numbered box. Some of the symbols in the column are not in the key. In this case, subjects should not use them in any of the boxes. Some of the numbers will not have corresponding symbols in the column. In this case, subjects should leave the boxes with those numbers empty. Each box, therefore, can contain either one or no symbols. Figure 2.1b shows an example of a correctly solved round. Note that boxes 3 and 5 are left empty, because there are no corresponding symbols, ð and ≈, in the column.

After filling in the boxes as subjects see fit, they click "Submit" to proceed to the next round. Subjects can complete each round at their own pace, however, they are asked to complete all 100 rounds of the task. Subjects learn their scores at the end of the experiment.

The cognitive task used in this experiment is a version of a neuropsychological test known as the digit symbol test (DST).[17] A typical version of DST presents subjects with a table of correspondences between digits from 1 to 9 and random symbols and asks subjects to fill in the blank spaces under numbers with a corresponding symbol. Subjects have a fixed amount of time (usually 90 or 120 seconds) to answer as many questions as possible. Versions of DST appear in the Wechsler Adult Intelligence Scale (WAIS), which is one of the most commonly used IQ tests in the world (Weiss, Saklofske, Coalson, and Raiford, 2010), in the ASVAB test battery in the NLSY panel, and in the Unified HuntingtonâĂŹs Disease Rating Scale (Kieburtz, 1996). Performance on DST is associated with such cognitive traits as processing speed and working memory. DST finds a widespread use in neuropsychological and medical literatures as a reliable measure of cognitive ability. It has been used to study patients with schizophrenia, who show lower scores and higher response times on the test (Amaresha, Danivas, Shivakumar, Agarwal, Kalmady, Narayanaswamy, and Venkatasubramanian, 2014), unsafe driving in old-aged patients

[17] Other names for DST are coding test, coding speed test, digit symbol substitution test, and symbol digit modalities test.

with and without Alzheimer disease, for whom DST is shown to be the best predictor of unsafe driving (Lafont, Marin-Lamellet, Paire-Ficout, Thomas-Anterion, Laurent, and Fabrigoule, 2010), concussions among athletes (Hunt and Asplund, 2010), and other deceases affecting cognitive function. In economics studies, researchers have used DST to study the relation between cognitive ability and risk and time preferences in incentivized experiments (Dohmen, Falk, Huffman, and Sunde, 2010) and the role of motivation in performance (Segal, 2012).

The benefit of a cognitive task like DST is that it does not rely on any special knowledge and thus measures fluid intelligence, as opposed to crystallized intelligence (Cattell, 1971). Fluid intelligence is responsible for solving novel problems that do not rely on any cultural background or accumulated knowledge for solution.[18] The measure of processing speed that DST provides is positively associated with other IQ measures, as the processing speed is the basis for other, more complex, cognitive functions (Vernon, 1983). DST compares favorably to other cognitive tests since it is able to detect small changes in cognitive ability among subjects with normal levels of cognition (Proust-Lima, Amieva, Dartigues, and Jacqmin-Gadda, 2007), such as those used in our study.

The version of DST used here differs from the canonical implementation in several ways. I do not fix the time to complete the task, but rather fix the number of trials. The reason for not imposing a time limit is to obtain the estimates of both ability and motivation. The proposed theoretical model assumes that response time is the main margin through which subjects affect their effort accumulation. It is necessary to ensure that response time is a choice variable. The raw score on the canonical DST—the amount of correct answers in a given period of time—provides a measure of speed, and it is related to the ability measure we are estimating—average number of correct trials per unit of time—but in addition to ability, I am also able to estimate motivation. It is possible, in principle, for subjects to complete each trial correctly if they are willing to spend enough time on carefully substituting the right numbers for symbols. Spending more time on

[18]For instance, other popular tasks such as an addition task, counting task, or a paragraph completion task do require such background and knowledge, and therefore are better suited for measuring crystallized intelligence. Other popular choices for cognitive tasks that rely on fluid intelligences are Raven's Matrices (Gill and Prowse, 2016) and a game of Tower-of-Hanoi (McDaniel and Rutström, 2001).

a trial pays out since it helps to eliminate possible mistakes. Mistakes can easily be made in our version of DST, since some symbols are not from the key, and some symbols look similar. Subjects who are not willing to spend much time on each trial are likely to have a less-than-perfect score. Lastly, the key is different in every trial, as I am primarily interested in using the test to gauge processing speed rather than working memory, which does contribute to the DST scores though less prominently than speed (Joy, Kaplan, and Fein, 2004).

Subjects have three practice rounds before the actual task begins. This gives them a chance to familiarize themselves with the interface. During the practice, they receive feedback if they make a mistake. Subjects receive no feedback on their performance between the actual rounds.[19] Practice trials ensure that performance on a task is not influenced by the confusion about the task interface. They also capture any possible learning effects so that the subsequent performance on a task does not reflect learning but only the parameters of interest: ability and motivation. In order to further prevent learning effects and encourage independent decisions in each trial, the feedback about performance in the non-practice rounds is provided only after the last round of the task.[20]

Subjects receive a flat payment of $20 for completing the cognitive task and receive no additional monetary reward for correctly solved trials. This allows me to measure subjects' intrinsic motivation. In the absence of intrinsic motivation, subjects should not put any effort in solving the task correctly and their responses should not be different from random guessing. The probability of success should be zero in this case since no answers to choose from are provided. On the flip side, if subjects are intrinsically motivated and are willing to put effort in solving the problems their probability of success would be significantly different from zero.[21]

[19] Studies have shown that quick feedback is crucial for learning and improving performance (Balzer, Doherty, and O'Connor, 1989; Hoch and Loewenstein, 1989).

[20] The feedback is not provided after each round in the present design, since it would likely make decisions in trials interdependent. For example, a subject who solved one trial incorrectly and knows about it might try harder next time to compensate for the mistake.

[21] It is also possible that non-zero probability of success can be driven by positive reciprocity towards an experimenter. However, it is clear from the instructions that the task does not benefit anyone outside the lab or the experimenter. Thus, the only plausible benefit from the correct solution seems to be the subjects' own satisfaction with the result, i.e., intrinsic motivation.

Risk task

The risk task follows the GARP-style design of Choi, Kariv, Müller, and Silverman (2014). The task asks subjects to allocate points between the two accounts. This choice is implemented graphically by selecting an allocation on the budget line (see Figure 2.2).[22] Each of the accounts is equally likely to be picked for payoff, in which case a subject receives the points allocated to this account. The allocations above the equal-allocation point introduce variation in possible payoffs, as they pay more in case y is picked and less in case x is picked. At the same time, they have higher expected values, as the budget line is steep in this example: one point in account x can be converted to two points in account y. The allocations below the equal-allocation point introduce variance and have smaller expected values, and thus are inferior to other allocations: subjects satisfying the first-order stochastic dominance should not select them. Subjects are given several trials of this task, which differ in the steepness and position of the budget lines presented. The budget lines are generated by randomly drawing the values for the maximum possible token allocation to accounts x and y from a uniform distribution with support $[50, 100]$. After a subject completes all the trials, a random trial is selected, and the paying account is randomly picked, which determines a subject's payoff for the task.

The benefit of this particular risk task is that it allows to gauge both a non-parametric measure of risk aversion (defined as the proportion of tokens allocated to a more expensive account relative to the risk-free allocation),[23] as well as a non-parametric measure of choice consistency (CCEI) due to Afriat (1972). I can then look at how my estimates of ability and motivation are related to these two measures. The relative effects of ability and motivation on consistency are particularly interesting, as Choi *et al.* (2014) find that consistency is positively associated with life-time earnings.

[22] The software for this task is provided courtesy of Shachar Kariv.

[23] This definition is slightly different from the one used in Choi *et al.* (2014). The measured used here is increasing in risk aversion and has a maximum of 1 and a minimum of 0, which makes it easier to interpret. Appendix B.2 describes in more detail how risk aversion is calculated and how this measure relates to a parametric estimate of the constant relative risk aversion.

Subjects participate in 25 rounds of this task. Only one of them determines the payment for this task. At the end of the experiment, subjects throw a die to pick one of the rounds at random (each round has an equal chance of being picked). After the round that counts is picked, subjects throw a die again to randomly pick account x or account y (each account has an equal chance of being picked) as the paying account. The number of tokens a subject allocated to the paying account determines their payment for the task. The tokens are converted into dollars at the rate of 5 tokens = $1.

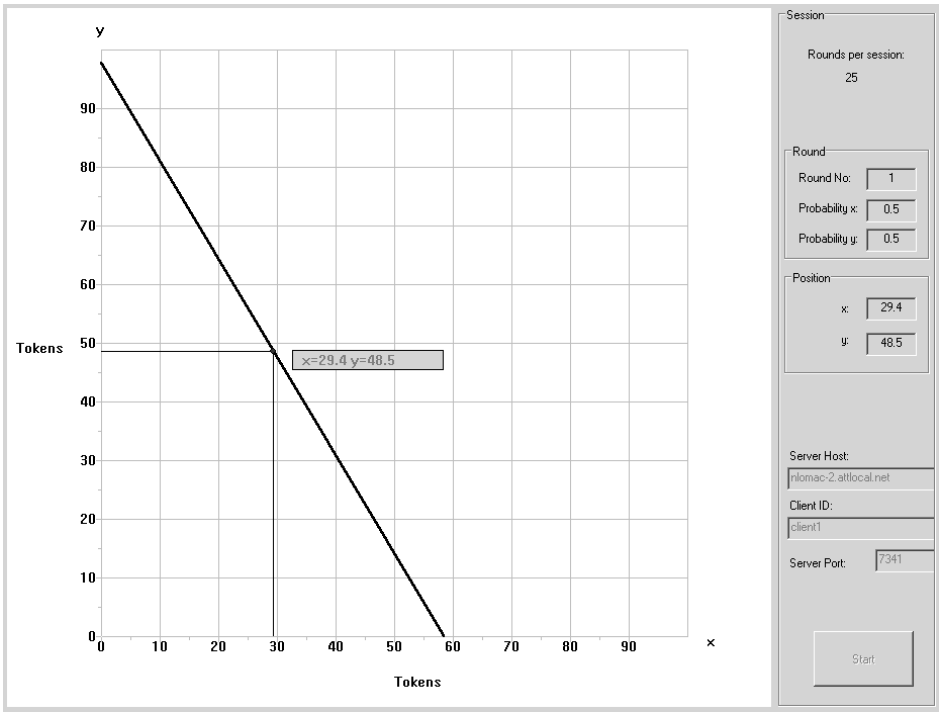


Figure 2.2: Risk Task, Decision Screen

Additional measures

After completing the two decision tasks, subjects were presented with the survey questions. The questionnaire included demographic questions, 3-question Cognitive Reflection Test (Frederick, 2005), and personality questions. CRT is frequently used as a measure of cognitive ability and it is interesting to compare it with the estimates of ability and motivation. The Big Five personality test is a popular tool for gauging personality traits. I use five facets from the Big Five that can be

directly related to my estimated measures of ability, motivation, and consistency: Activity level (e.g., "Do a lot in my spare time", "Like to take my time"), Intellect (e.g., "Like to solve complex problems", "Am not interested in abstract ideas"), Self-efficacy (e.g., "Excel in what I do", "Have little to contribute"), Achievement-striving (e.g., "Plunge into tasks with all my heart", "Am not highly motivated to succeed"), and Cautiousness (e.g., "Choose my words with care", "Do crazy things").

2.5 Results

2.5.1 Ability

I first present the results for the estimated ability. Figure 2.3 (left panel) shows the histogram and the kernel density estimate of the distribution of ability in the sample, with the dashed line showing the median and the dotted line showing the reference density of a normal distribution with the mean and variance equal to the corresponding sample moments of ability. The median ability is 7.3, minimum ability is 2.3, and maximum ability is 14.7. The distribution of ability has a bell shape and is well concentrated around the median. The distribution is far from being normal, however, due to fat right tail and a positive skew (Shapiro-Wilk test p-value < 0.001). Relative to a normal distribution with the same sample mean and variance, the distribution of ability has a higher mass of subjects with ability slightly below the average and with very high ability. On the other hand, the distribution of ability has a lower mass of subjects with very low ability and ability slightly above the average.

The ability parameter has a clear meaning of the speed of effort accumulation, however it is more intuitive to interpret its magnitude through its effect on the probability of success. I use formula (2.4) to compute the optimal effort threshold and implied probability of success for each ability value in the sample for fixed values of motivation and consistency.[24] I use the median of

[24] An alternative way of interpreting the values of ability is to look at the average probabilities of success that each subject would reach after t seconds, where t could be set, e.g., such that a subject with a median ability reaches a 0.5

motivation and consistency as such fixed values. The resulting quantity yields the *counterfactual* distribution of the probabilities of success that would arise from the variation in ability alone and is stripped down from all the variation in motivation and consistency. Figure 2.3 (right panel) shows the kernel density estimate of the distribution of the resulting implied probabilities of success and, for reference, the distribution of the actual probabilities of success implied by the raw scores on the cognitive task.

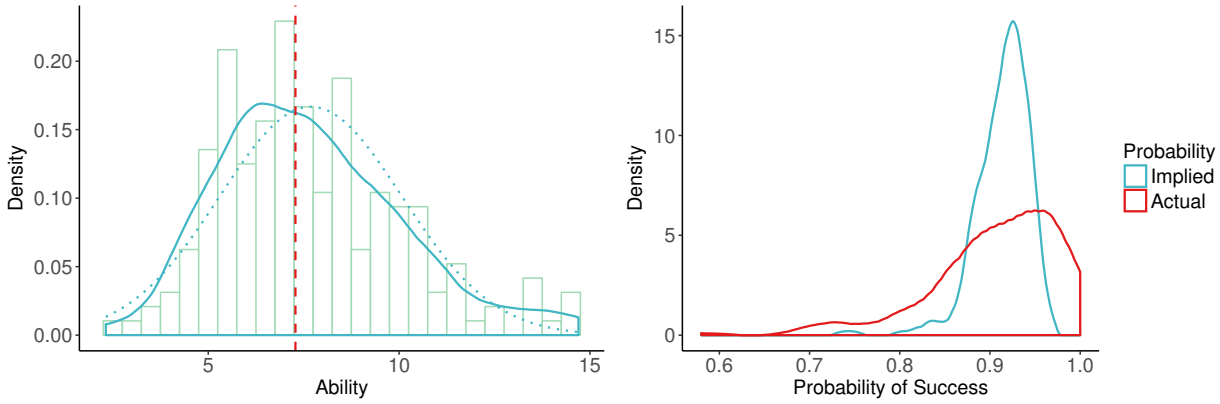


Figure 2.3: Distribution of Ability and Probability of Success

The graph reveals a striking difference between the distributions of the raw probabilities of success and the probabilities of success implied by ability alone.[25] The distribution of implied probabilities is much tighter than the distribution of actual probabilities, which should not come as a surprise, since the former is filtered out of all the variation in motivation and consistency.[26] The medians of implied and actual probabilities are almost the same, around 0.92, but the standard deviation of the actual probabilities is more than twice as high (0.07 versus 0.03). There is a high degree of association between the actual and implied probabilities, as one would expect, with Kendall's $\tau = 0.54$ (p-value < 0.001), however, the two distributions are clearly not iden-

probability of success in t seconds. Clearly, subjects with higher ability would reach higher probabilities of success in a given amount of time, on average, than subjects with lower ability.

[25]Using means of motivation and consistency as fixed values instead of medians leads to similar conclusions. In fact, the counterfactual distribution in this case is even tighter. See Figure B.4 (left panel) in Appendix B.5.

[26]The two distributions would be similar if, e.g., there were little variation in motivation among subjects, which apparently is not the case here.

tical (Kolmogorov-Smirnov test p-value < 0.001).[27] This illustrates the drawback of using a raw performance score as a proxy for ability, since it results in a much noisier estimate.[28]

The distribution of implied PoS shows that there is substantial variation in ability among subjects, even though this variation is not as large as suggested by a noisy raw measure. To gauge the magnitude of this variation, Table 2.2 shows two measures of "inequality" among the subjects in the sample by their ability. The first measure shows the average implied PoS for different slices of the sample. Moving from the bottom 1% to the top 1% by ability (in Panel A) results in an increase in the average implied PoS from 0.78 to 0.96, or by 24%. Moving from the bottom 10% to the top 10% results in an increase in the average implied PoS from 0.86 to 0.95, or by 11%. Finally, moving from the bottom 50% to the top 50% results in a modest increase in the average implied PoS from 0.89 to 0.94, or by 5%.

The second measure is a "share of income" type measure that shows a share of income for each group, if that income was allocated based on subjects' productivity as measured by their PoS. It can be viewed as a renormalized mean PoS measure. For example, the share of the bottom 1% is 0.009 while the share of the top 1% is 0.011. Even in these two extreme groups the inequality in income caused by the differences in ability is not excessive. This means that despite there is a number of subjects with very high ability, this does not translate into excessively high PoS.

The inequality between groups due to differences in ability is mediated by the level of motivation at which the implied PoS is computed. In particular, low motivation, defined as an average motivation for the bottom 10% of subjects by motivation, leads to a greater inequality due to differences in ability, as shown in Panel B of Table 2.2. In this case, moving from the bottom 1% to the top 1% by ability results in an increase in the average implied PoS from 0.55 to 0.92, or by 68%. Low motivation leads to higher shares of each "top" group and lower shares of each "bottom"

[27] All the two-samples tests are two-sided, unless otherwise noted.

[28] The fact that the distribution of implied PoS is much tighter than the distribution of actual PoS does not mean that the implied PoS ignores individual heterogeneity. On the contrary, it makes studying heterogeneity in ability more informative, since there is no confounding variation in motivation or consistency that is present in a raw measure of performance. For example, suppose that males have higher ability than females, while females have higher motivation than males. Using performance as a measure of ability might lead to a wrong conclusion that there are no gender differences in ability since the two gender effects cancel out. Using the proposed measure, on the other hand, will allow one to correctly pick up the differential gender effects on ability and motivation.

group relative to the case of median motivation. On the other hand, high motivation, defined as an average motivation of the top 10% of subjects by motivation, leads to lower inequality. For high motivation, moving from the bottom 1% to the top 1% by ability results in an increase in the average implied PoS from 0.88 to 0.98, or only by 11%. High motivation leads to lower shares of each "top" group and higher shares of each "bottom" group relative to the case of median motivation.

Table 2.2: Measures of Inequality Between Groups by Ability

Measure	Bottom 1%	Bottom 10%	Bottom 50%	Top 50%	Top 10%	Top 1%
Panel A. *Median Motivation*						
Mean PoS	0.775	0.856	0.893	0.935	0.952	0.959
Share	0.009	0.098	0.488	0.512	0.108	0.011
Panel B. *Low Motivation*						
Mean PoS	0.546	0.710	0.784	0.870	0.902	0.917
Share	0.007	0.089	0.474	0.526	0.114	0.012
Panel C. *High Motivation*						
Mean PoS	0.880	0.923	0.943	0.965	0.974	0.978
Share	0.010	0.101	0.494	0.506	0.106	0.011

2.5.2 Motivation

I next present the results for the estimated motivation. Figure 2.4 (left panel) shows the histogram and the kernel density estimate of the distribution of motivation in the sample, with the dashed line showing the median and the dotted line showing the reference density of a normal distribution with the mean and variance equal to the corresponding sample moments of motivation. The median motivation is 1.7, minimum motivation is 0.6, and maximum motivation is 20.5. The distribution of motivation has a bell shape but is heavily positively skewed due to a very long right tail. Relative to a normal distribution with the same sample mean and variance, the distribution of motivation has a much larger mass of subjects with motivation below the average and with very high motivation. However, the distribution of motivation has a much lower mass of subjects with motivation slightly above the average.

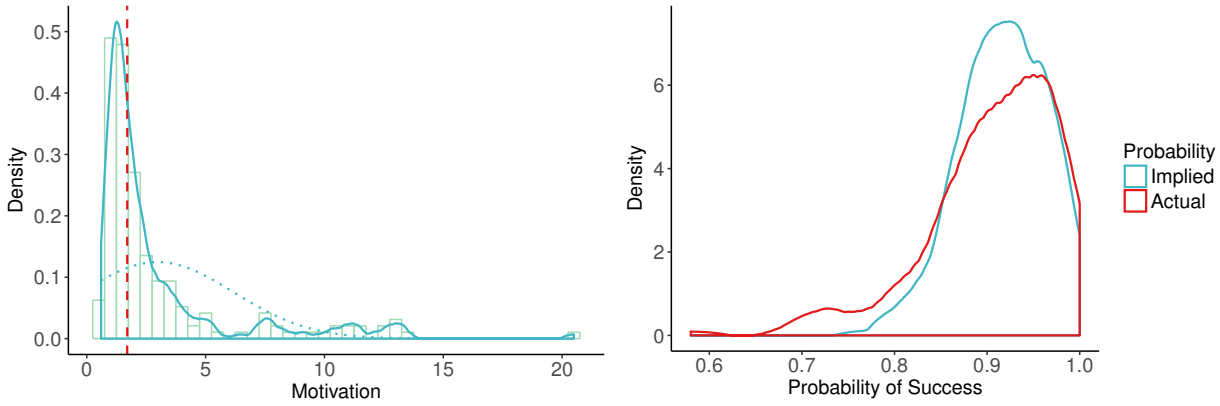


Figure 2.4: Distribution of Motivation and Probability of Success

The magnitude and variation in motivation is most transparent when one looks at its effect on the PoS. Using equation (2.4), I simulate the counterfactual distribution of the PoS due to variation in motivation, with the ability and consistency channels shut down by setting them at a median level. Figure 2.4 (right panel) shows the resulting implied distribution of the PoS, along with the distribution of the actual PoS. The two distributions look remarkably similar.[29] They have roughly the same median of 0.92, but the distribution of the implied PoS has a smaller standard deviation of 0.05 (versus 0.07 for the distribution of the actual PoS). Smaller standard deviation is the consequence of eliminating the variation in ability and consistency that is present in the actual PoS. The implied and actual PoS have a very high degree of association, Kendall's $\tau = 0.79$ (p-value < 0.001). In fact, one cannot reject at a 5% level the null hypothesis that the two samples come from the same underlying distribution (Kolmogorov-Smirnov test p-value = 0.127). This suggests that the observed variation in the actual PoS is mostly caused by the variation in motivation, rather than variation in ability.

Table 2.3 presents the measures of inequality due to differences in motivation among different quantiles of subjects by their motivation. Consistent with Figure 2.4 (right panel), variation in motivation induces a much greater variation, and hence inequality, in terms of PoS among subjects than ability. Panel A shows the results for the case when the implied PoS is computed at the

[29]Using means of ability and consistency leaves the results virtually unchanged. See Figure B.4 (right panel) in Appendix B.5.

median ability. Moving from the bottom 1% to the top 1% by motivation results in an increase in the average implied PoS from 0.79 to 0.99, or by 26%. Moving from the bottom 10% to the top 10% results in an increase in the average implied PoS from 0.83 to 0.99, or by 19%. Finally, moving from the bottom 50% to the top 50% results in an increase in the average implied PoS from 0.88 to 0.95, or by 9%. Moving across different groups my motivation, say, from bottom 10% to top 10%, results in a higher impact on the implied PoS than for similar groups by ability.

Greater inequality due to variation in motivation can also be observed for the share measure. Comparing shares across Tables 2.2 and 2.3 reveals that each "top" group by motivation has a higher share than the corresponding "top" group by ability. For example in Panels A, the share of the top 10% by motivation is 0.112, while the share of the top 10% by ability is only 0.108. Similarly, each "bottom" group by motivation has a lower share than the corresponding "bottom" group by ability.

As was the case with ability, the inequality measures in terms of the implied PoS due to variation in motivation are mediated by the value of ability taken as a reference. Using high ability, defined as an average ability for the top 10% of subjects by ability, as a fixed value results in greater inequality among groups of subjects relative to the case of median ability, see Panel B of Table 2.3. On the other hand, using low ability, defined as an average ability for the bottom 10% of subjects by ability, as a fixed value results in lower inequality among groups of subjects relative to the case of median ability, see Panel C of Table 2.3.

2.5.3 Relation Between Measures

Table 2.4 reports the rank correlations (Kendall's τ) between estimated measures of ability, motivation, and consistency. It indicates a significant positive association between ability and motivation, but a negative association between ability and consistency. This suggests that subjects who have higher ability are, on average, more motivated but are also less consistent. The negative association between motivation and consistency is rather strong and suggests that highly motivated subjects were also among less consistent.

Table 2.3: Measures of Inequality Between Groups by Motivation

Measure	Bottom 1%	Bottom 10%	Bottom 50%	Top 50%	Top 10%	Top 1%
Panel A. *Median Ability*						
Mean PoS	0.787	0.831	0.879	0.955	0.986	0.990
Share	0.009	0.094	0.479	0.521	0.112	0.011
Panel B. *Low Ability*						
Mean PoS	0.640	0.715	0.796	0.924	0.976	0.983
Share	0.008	0.087	0.463	0.537	0.118	0.012
Panel C. *High Ability*						
Mean PoS	0.875	0.901	0.930	0.974	0.992	0.994
Share	0.010	0.099	0.488	0.512	0.109	0.011

Table 2.4: Rank Correlation Between Measures

	Ability	Motivation	Consistency
Ability	1.000	0.303***	−0.205***
Motivation		1.000	−0.512***
Consistency			1.000

Notes: Kendall's rank correlation coefficients reported, *** denotes significance at a 0.1% level.

2.5.4 Determinants of Ability

In this part, I consider whether the observed variation in estimated ability can be explained by the observed variation in demographic characteristics, preferences, and character skills. To facilitate the interpretation of results, instead of focusing on raw estimates of ability, I again use the PoS implied by a given level of ability (ability-PoS), with motivation and consistency set at median levels. Given a strictly monotonic relationship between ability and PoS implied by it, I will use the terms ability and implied PoS interchangeably in this section.

Figure 2.5 shows the kernel density estimate (left panel) and empirical CDF (right panel) of the counterfactual distributions of the implied PoS by gender. The graphs suggest that females tend to have higher ability, and hence higher implied PoS, than males. The median implied PoS for females is 0.92, which is 1% higher than the median implied PoS for males, 0.91. The difference between males and females is small, but statistically significant (Wilcoxon Rank Sum test p-value = 0.004).[30] The ability differences between males and females become more pronounced at lower quantiles. The distribution of the implied PoS for females is less dispersed, it has a lower standard deviation than that for males (0.02 for females versus 0.04 for males).

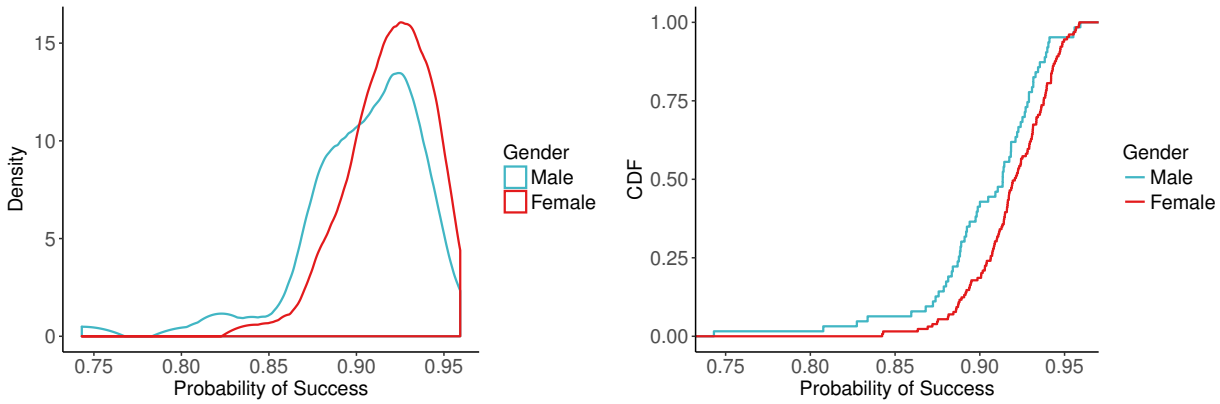


Figure 2.5: Distribution of Ability-PoS by Gender

Figure 2.6 shows the kernel density estimate (left panel) and empirical CDF (right panel) of the counterfactual distributions of the implied PoS by race. The median implied PoS is 0.91 for Black subjects, 0.92 for White subjects, and 0.93 for Asian subjects. Black subjects tend to have slightly

[30]The results of the two-sample tests in this section are robust to using Kolmogorov-Smirnov test, instead.

lower implied PoS than White (Wilcoxon Rank Sum test p-value = 0.075) or Asian (Wilcoxon Rank Sum test p-value = 0.006) subjects. The implied PoS for Asian and White subjects, on the other hand, does not differ significantly (Wilcoxon Rank Sum test p-value = 0.448). White subjects tend to have higher implied PoS than Asian subjects for lower quantiles, but the pattern is reversed for higher quantiles. White subjects also have the lowest dispersion by ability, with a standard deviation of 0.02.

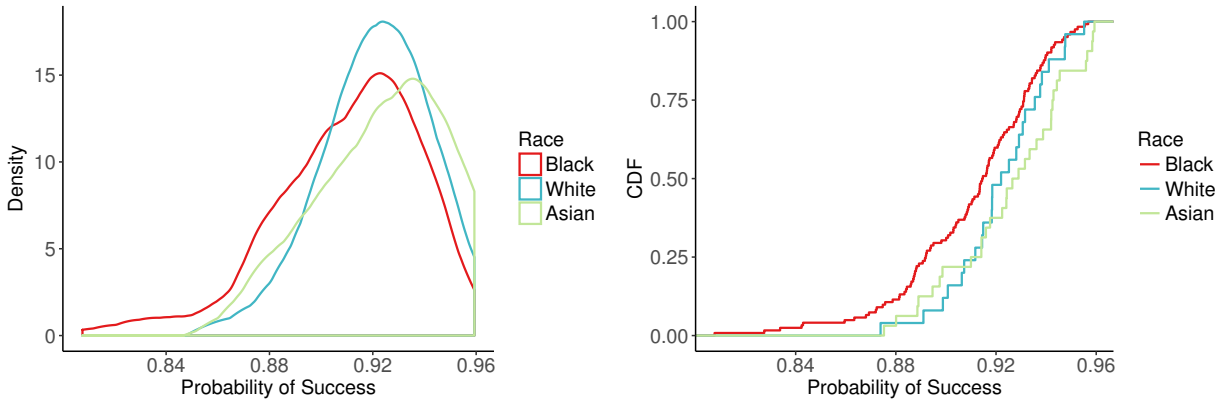


Figure 2.6: Distribution of Ability-PoS by Race

Table 2.5 (columns 1-3) reports the results of a fractional regression of the implied PoS on a set of demographic, cognitive, preference, and character covariates. Column 1 shows that the gender and race effects are robust to inclusion of additional demographic controls, such as major, year in school, age, religion, family income, awards, work, and spending.[31] However as column 2 shows, the race effects become muted after controlling for subjects' cognitive abilities using the traditional measures of CRT score and self-reported GPA. This suggests that the racial differences observed in specification 1 mostly reflect differences in cognitive abilities.

Specification 3 explores the relation between the implied PoS and cognitive abilities, risk preferences (risk aversion and CCEI), and character skills (facets from the factors in the Big Five). Results show that both CRT and GPA are positively related to the implied PoS, with GPA being statistically significant at the 10 percent level. The CCEI measure from the risk task is highly

[31] Additional controls are jointly insignificant and thus are excluded from other specifications. The only effect close to being significant is that subjects who report medium level of family income tend to have lower ability than subjects who report either low or high family income.

significant, suggesting that the estimated measure of ability is related to the quality of decision-making. This result is consistent with the findings in Choi, Kariv, Müller, and Silverman (2014) about the positive association between CCEI, as a measure of decision-making quality, and cognitive ability. Risk aversion is positively related to the implied PoS, however, the effect is imprecisely estimated. The positive association between ability and risk aversion, however, is not consistent with the findings in Dohmen, Falk, Huffman, and Sunde (2010).[32]

Most surprisingly, I do not find any association between any of the measures of character skills and implied PoS. The three measures considered, Intellect, Activity Level, and Self-Efficacy, were selected because of their conceptual similarity to the introduced here model-based ability parameter. It is natural to expect that these self-reported measures would be, at least weakly, positively associated with the objectively measured ability. In reality, all of the self-reported measures have small and insignificant coefficients, and two of the three measures have negative coefficients. Two possible explanations arise. First, the self-reported characters skills are accurate and measure general tendencies, while the objectively measured ability in the cognitive task is very context-dependent. This explanation seems unlikely, since it implies that the general ability, as measured by the self-reports, has no (or even negative) association with the objectively measured ability in a well-established cognitive task.

An alternative explanation is that the self-reported measures are biased to the point that they distort the relation between the objective and self-reported measures. It is well-known that people tend to be overconfident in their abilities relative to others, for example, due to self-image concerns. Similarly, subjects are found to misreport information when it is beneficial to do so in the experiments on lying. It is plausible, therefore, that when asked about a socially desirable skill, such as intellect or ability, an individual has an incentive to artificially boost his or her report about that skill. If this effect is stronger for subjects whose true ability is low, the self-reported measure

[32]This could be caused by the differences in the risk elicitation procedures. Dohmen *et al.* (2010) used the MPL procedure, while I am following the GARP-style design of Choi, Fisman, Gale, and Kariv (2007). The unconditional rank-order correlation between ability and risk aversion is negative but not statistically significant. I also repeat the analysis for a subset of subjects for whom risk aversion is not too high (< 0.95) in Table B.1. The results are qualitatively similar.

will completely distort the positive association between the report and the truth. This suggests that the answers to the survey questions about desirable character skills might not be reliable.

Table 2.5: Fractional Regression Results

	(1)	(2)	(3)	(4)	(5)	(6)
Constant	2.120***	2.030***	1.090**	2.260***	2.040***	2.060**
	(0.141)	(0.189)	(0.507)	(0.258)	(0.330)	(0.848)
Demographics						
Female	0.197***	0.208***		0.128	0.144	
	(0.054)	(0.053)		(0.099)	(0.094)	
White	0.181**	0.123		0.001	−0.117	
	(0.086)	(0.082)		(0.153)	(0.140)	
Asian	0.216***	0.146*		−0.084	−0.264**	
	(0.080)	(0.077)		(0.136)	(0.125)	
Cognitive Abilities						
CRT		0.039	0.031		0.114**	0.072
		(0.026)	(0.028)		(0.046)	(0.047)
GPA		0.042	0.112*		0.075	0.052
		(0.059)	(0.061)		(0.103)	(0.101)
Preferences						
Risk Aversion			0.087			0.196
			(0.132)			(0.212)
CCEI			1.020***			0.401
			(0.389)			(0.671)
Character Skills						
Intellect			0.001			
			(0.005)			
Activity Level			−0.005			−0.015
			(0.006)			(0.010)
Self-Efficacy			−0.002			
			(0.006)			
Achievement Striving						0.001
						(0.009)
Additional Controls	Yes	No	No	Yes	No	No
Observations	192	192	192	192	192	192

Notes: Reports the coefficients of a fractional regression. Models 1-3 use ability-PoS as a dependent variable, models 4-6 use motivation-PoS as a dependent variable. Standard errors in parenthesis. Omitted categories are female and Black (African American). Additional controls include major, year in school, age, religion, family income, whether a student receives an award, whether a student works, and weekly spending.

*** Significant at the 1 percent level.

** Signficant at the 5 percent level.

* Significant at the 10 percent level.

2.5.5 Determinants of Motivation

In this part, I turn to motivation and ask whether the observed variation in it can be explained by the observed variation in demographic characteristics, preferences, and character skills. Once again, to facilitate the interpretation of results, I use the PoS implied by a given level of motivation, with ability and consistency set at median levels. The terms motivation and implied PoS (or motivation-PoS) will be used interchangeably in this section.

Figure 2.7 shows the kernel density estimate (left panel) and empirical CDF (right panel) of the counterfactual distributions of the implied PoS by gender. The graphs provide little evidence for gender differences in motivation. Females tend to have higher implied PoS for some quantiles, but overall the distributions for females and males look similar. The median implied PoS for males and females is around 0.92. The difference in motivation between males and females is not statistically significant (Wilcoxon Rank Sum test p-value = 0.249).

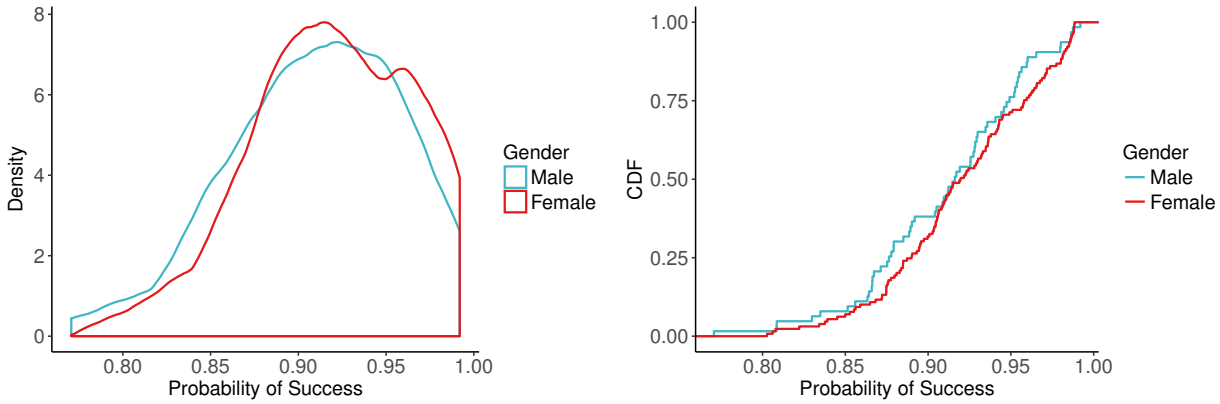


Figure 2.7: Distribution of Motivation-PoS by Gender

Figure 2.8 shows the kernel density estimate (left panel) and empirical CDF (right panel) of the counterfactual distributions of the implied PoS by race. The graphs show no discernible race effects for motivation. The median implied PoS is around 0.92 for all races. The differences in motivation between races are not statistically significant (Wilcoxon Rank Sum test p-values are = 0.871 for Black versus white, = 0.537 for asian versus black, and = 0.744 for asian versus white). Asian subjects tend to have lower implied PoS than Black or White subjects for lower quantiles,

but the pattern disappears for higher quantiles. Asian subjects have a slightly higher dispersion by motivation than Black or White subjects, with a standard deviation of 0.06.

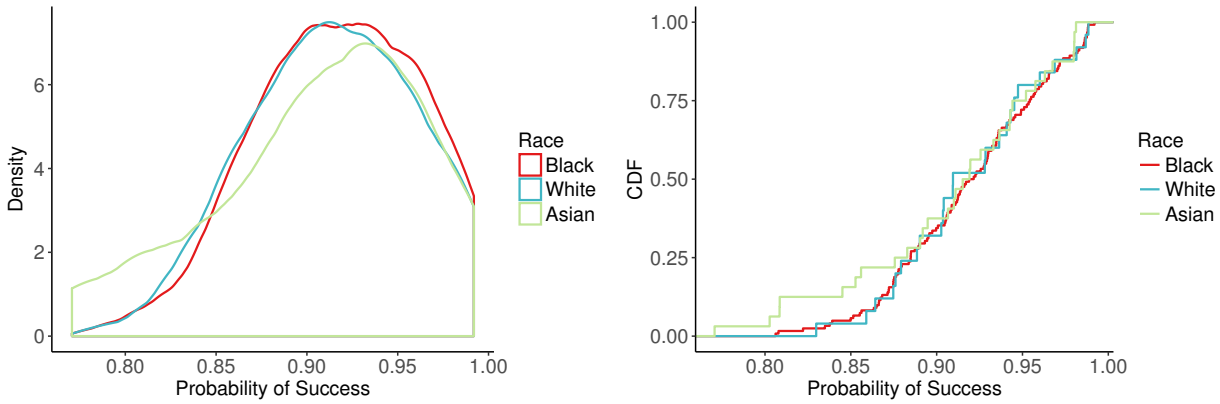


Figure 2.8: Distribution of Motivation-PoS by Race

Table 2.5 (columns 4-6) reports the results of a fractional regression of the implied PoS (or motivation-PoS) on a set of demographic, cognitive, preference, and character covariates. Column 4 confirms that there is little discernible association between gender, race, and motivation, even after controlling for additional demographic variables.[33] Females have higher motivation than males, but the effect is much smaller relative to that in case of ability and is imprecisely estimated. The difference in motivation between black and white subjects is virtually zero, though the standard error on the indicator for white subjects is high. Asian subjects tend to have lower motivation than black subjects, but the effect is imprecisely estimated. Specification 5 uses cognitive measures as additional controls. Using these additional controls makes the negative coefficient on the indicator for Asian subjects statistically significant at the 5% level. Out of the two cognitive measures used, only CRT score is statistically significant.

Specification 6 explores the relationship between motivation and cognitive measures, preferences, and character skills. None of the cognitive measures appears to be in a significant association with motivation, even though all of them have positive coefficients. Similarly, the effects of risk aversion and CCEI are very imprecisely estimated. The coefficient on risk aversion is positive,

[33] As was the case with ability, additional controls are jointly insignificant and thus are excluded from other specifications.

suggesting that subjects who are more motivated are also more risk averse.[34] Overall, it appears that motivation is a much more idiosyncratic characteristic of an individual than ability, given a weak association between the implied PoS and various predictors.

Interestingly, the self-reported measure of motivation (Achievement Striving) has virtually no association with the objective measure of motivation. The coefficient on Achievement striving, while positive, is very small and has a large standard error. The effect of self-reported Activity Level, which could also be broadly related to motivation, is very small and imprecisely estimated. As was the case with ability, the self-reported and objective measures of motivation are virtually unrelated. This raises concerns about the validity of the self-reported measures, since it is natural to expect that the true general tendency to be highly motivated should be related to the objective measure, even elicited in a specific context. It is possible, of course, that the objective measure of motivation elicited in the cognitive task is context-specific and thus represents a completely different characteristics of an individual, but this explanation seems unlikely. It appears more plausible that subjects inflate their reports of motivation in the questionnaire, since high motivation is a socially desirable characteristic. If this effect is strong enough among low-motivation subjects, this could drive the relationship between the report and the truth to zero.[35]

The present results, while suggestive of the potential issues with relying on self-reported measures of cognitive or character skills, are far from being conclusive. More research is required on the relationship between self-reported and objective measures of ability and motivation. One line of research should consider alternative cognitive tasks to evaluate the sensitivity of this relationship to various contexts. Another important direction of further research is to use the objective and self-reported measures as explanatory variables for relevant outcomes, such as life-time income, risky behavior, or the quality of financial decisions, similarly to what has been done by, e.g., Noussair,

[34] This result is not robust, however, if one considers a subset of subjects with a moderate degree of risk aversion, see Table B.1. In fact, the effect of risk aversion has the opposite sign in that case.

[35] It would be interesting to consider an association between the proposed measure of motivation and the Big Five factor of conscientiousness, since it is sometimes described as the propensity to work hard, persevere, and strive for success (Barrick and Mount, 1991). In fact, the achievement-striving measure used here is a part of a conscientiousness factor. I opted for using the more narrowly defined achievement-striving measure since its meaning directly corresponds to the meaning of the proposed measure of motivation, as opposed to conscientiousness, which is a composite trait.

Trautmann, and van de Kuilen (2014) or Choi, Kariv, Müller, and Silverman (2014) in the context of risk attitudes.

2.5.6 Ability, Motivation, and Success

In this part, I investigate which of the subjects' characteristics, ability or motivation, has a higher impact on success in the cognitive task. I start by splitting the sample of subjects into groups defined by two binary characteristics: ability (high or low) and motivation (high or low).[36] This procedure yields four groups of subjects: group 1 (low ability, low motivation, LA-LM), group 2 (low ability, high motivation, LA-HM), group 3 (high ability, low motivation, HA-LM), and group 4 (high ability, high motivation, HA-HM). Figure 2.9 shows the kernel density estimates (left panel) and CDF (right panel) of the actual probabilities of success by group. The median PoS in each group are 86 (group 1), 93 (group 2), 91 (group 3), and 97 (group 4).

Several distinct patterns immediately arise from these graphs. Subjects with high ability have much tighter distributions than subjects with low ability. Keeping one characteristic constant and improving another characteristic always results in a significant increase in the PoS. For example, consider groups 2 and 1, both groups have low ability but group 2 has higher motivation. Higher motivation results in significantly higher values of the PoS for group 2 relative to group 1 (Wilcoxon Rank Sum test p-value < 0.001). Similarly, consider groups 3 and 1. Both groups have low motivation, but group 3 has higher ability. This results in significantly higher PoS for group 3 relative to group 1 (Wilcoxon Rank Sum test p-value < 0.001). Subjects with high values of both ability and motivation clearly dominate all other groups.

The question, however, is whether group 3 or 2, in which one characteristic is low and another is high, has higher PoS. If group 3 had higher PoS than group 2, this would imply that ability is a more important characteristic than motivation for achieving success in the cognitive task. If this were the case, starting in group 1 and moving to group 3 (i.e., improving ability) would have a higher impact on the PoS than moving to group 2 (i.e., improving motivation). The graphs suggest

[36] High (low) is defined by whether a characteristic is above (below) the median level.

otherwise. The right panel shows that group 2 stochastically dominates group 3 by a small margin. The left panel shows that this happens because in group 2 there is a non-trivial number of subjects reaching near-perfect score, while there are none of such subjects in group 3. The difference between groups 3 and 2, while not huge, is statistically significant (Wilcoxon Rank Sum test p-value = 0.007). Therefore, motivation turns out to be a more important characteristic than ability for achieving success.[37]

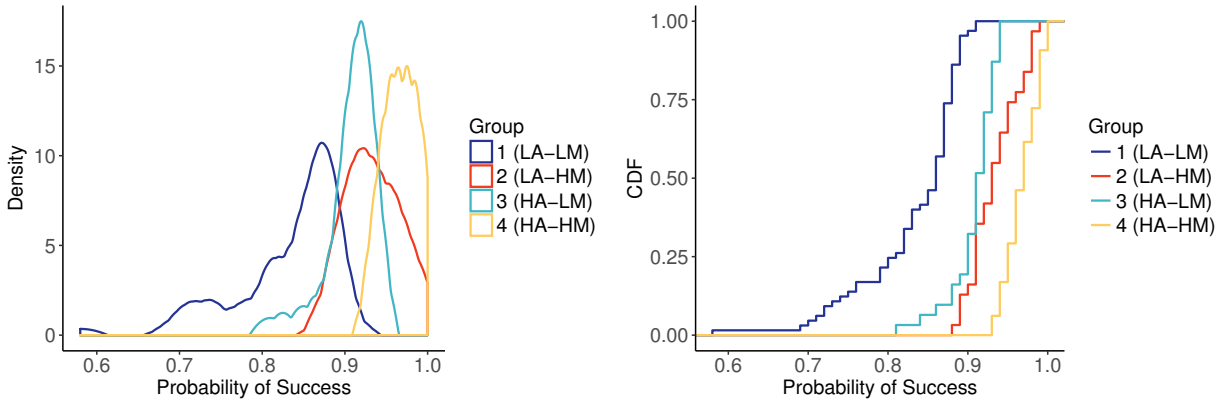


Figure 2.9: Distribution of PoS by Group

The result that motivation, a character skill, matters more for success than cognitive ability is consistent with the findings by Chetty, Friedman, Hilger, Saez, Schanzenbach, and Yagan (2011) who find evidence that improvement in character skills early in life has a long-lasting impact on future earnings, by Heckman, Pinto, and Savelyev (2013) who find that induced changes in character skills explain a larger proportion of adult outcomes than induced changes in cognitive ability, and by Nilsson (2017) who finds that character skills play a bigger role in explaining labor market outcomes. On the other hand, some studies find that cognitive ability matters more than character skills. For example, Gill and Prowse (2016) report that cognitive ability has a higher effect on the quality of strategic decision-making than agreeableness and emotional stability and that other character skills, such as conscientiousness or openness, have no effect. Nilsson (2017)

[37]The same conclusion can be reached by running a fractional regression of the actual PoS on the indicator variables for high ability and high motivation. The interaction term between ability and motivation has a significant negative coefficient. This implies that high ability and high motivation are substitutes: increasing motivation reduces the positive effect of ability, and *vice versa*.

shows that cognitive ability plays a bigger role in explaining educational outcomes. Choi, Kariv, Müller, and Silverman (2014) do not find a significant effect of conscientiousness on subjects' wealth. Overall, it appears that the importance of different skills varies by context and no single skill dominates others across the board.

2.6 Conclusion

The relative role of cognitive ability and character skills in life is an important question in the economics of human development. Existing methods of measuring cognitive ability and character skills suffer from two main flaws. First, the standard approach confounds ability with the combination of ability, motivation, and potentially other character skills, resulting in incorrect interpretations of the role of ability. Second, using self-reported statements for gauging character skills is problematic, since it is hard to assess their validity. They could represent wishful thinking rather than true characteristics. Additionally, self-reports about desirable social characteristics may be intentionally distorted because of the self-image concerns.

I propose an alternative method of measuring cognitive ability and motivation. My method improves upon the existing ones by *a*) decomposing the effects of ability and motivation on test scores and *b*) providing an objective measure of motivation. My method is based on a dynamic stochastic model of optimal effort choice with ability and motivation being the structural parameters of the model. I use response times as proxy for subjects effort, which links my approach to the literature on sequential-sampling, or drift-diffusion, models.

To assess the empirical validity of my method, I conduct a laboratory experiment in which subjects go through a series of trials of a Digit-Symbol test. I observe the outcomes of trials and response times, which together with a structural model allows me to uncover subjects' ability and motivation parameters. I also measure subjects' performance on a CRT, Big Five personality scores, risk preferences, the decision-making quality, and time preferences.

I find substantial heterogeneity among subjects in terms of their ability and motivation. Test scores turns out to be a very noisy measure of true ability. The observed variation in test scores is mostly due to variation in motivation. I find that females have higher ability than males, and that White and Asian subjects have higher ability than Black subjects. Subjects' consistency with utility-maximization in the risk task is strongly positively associated with higher ability. There is evidence that subjects who are more risk averse tend to have higher ability, but the effects of preferences are imprecisely estimated. Motivation turns out to be a more idiosyncratic characteristic than ability, as I do not find strong associations between motivation and gender, race, and other control variables.

I find virtually no association between estimated measures of ability and motivation and their self-reported counterparts from the Big Five questionnaire. The effects of the self-reported measures are close to zero and are very imprecisely estimated. I argue that self-reported measures may be biased to the point that they distort the true relation between objective and self-reported measures. This suggests that the answers to the survey questions about desirable character skills may be unreliable.

Looking at the relative importance of ability versus motivation on the success on a cognitive task, I find that motivation plays a slightly bigger role than ability. In particular, subjects with high motivation and low ability tend to have higher scores than subjects with low motivation and high ability. This pattern is driven by the fact that highly motivated subjects often reach near-perfect scores, while high-ability (but not motivated) subjects never reach such scores.

Chapter 3

Deciphering the Noise: The Welfare Costs of Noisy Behavior[1]

3.1 Introduction

Stochastic choice has become an active area of research in recent years. Developments in this area are motivated primarily by two considerations. First, a large body of empirical evidence shows that stochastic choice is a robust empirical phenomenon,[2] and much work has been devoted to explaining this behavior.[3] Second, models of stochastic choice provide researchers with econometric tools to estimate structural models in a broad range of applications. The primary interest in applying a model of stochastic choice is to recover the structural parameters of the deterministic part of a model, such as risk or time preferences. Little attention has been given, however, to the systematic economic interpretation of the parameter estimates of the stochastic part, which determine the magnitude of choice randomness. The interpretation of these parameters is important for understanding the economic value of choice randomness, which has implications for the

[1] Joint with Glenn W. Harrison, Morten Lau, and Don Ross.

[2] Nogee and Mosteller (1951) provide the earliest evidence of stochastic choice, followed by Tversky (1969), Starmer and Sugden (1989), Camerer (1989), and Ballinger and Wilcox (1997).

[3] Wilcox (2008) provides an excellent overview of many popular stochastic models of choice under risk. Recent examples include Swait and Marley (2013), Wallin, Swait, and Marley (2017), Matêjka and McKay (2015) and Agranov and Ortoleva (2017).

quality of decision making, and also for a better understanding of the underlying "source" models of stochastic choice. We study the economic consequences of stochastic choice and develop an intuitive method to translate the estimates of the stochastic part into economically tractable terms.

Consider a generic structural model of discrete choice that uses a standard multinomial logit model of stochastic choice[4], which assigns each discrete alternative a choice likelihood $\mathbb{P}$ according to

$$\mathbb{P}(a \mid \beta, \mu) = \frac{\exp(U(a \mid \beta)/\mu)}{\sum_{a' \in A} \exp(U(a' \mid \beta)/\mu)}. \tag{3.1}$$

In this expression, a and a' are alternatives, such as lotteries or dated outcomes, from a set of all alternatives A. The deterministic part of this structural model is parametrized by a vector of behavioral parameters β, which could represent, for instance, an agent's risk or time preferences. For example, in the case of risk preferences, β could be a risk aversion parameter and U could be the expected utility of a risky alternative; in the case of time preferences, such as the quasi-hyperbolic discounting model, β would comprise the exponential and hyperbolic discounting parameters and U would be the discounted utility of an income stream. The behavioral parameters determine the utility function U assigned to each alternative.

The stochastic part of the model is parametrized by μ, often called the noise parameter, which determines how sensitive choice likelihoods are to the maximization of utility according to a given structural model. In the extreme case of noise going to zero, the agent will almost surely choose the alternative with the highest utility. On the other end of the spectrum, when noise goes to infinity, the agent will assign equal likelihoods to choosing each alternative regardless of their utilities. Higher values of μ thus imply a higher magnitude of choice randomness in this popular specification.

Three issues arise with the interpretation of the noise parameter. First, while the effect of μ on choice likelihoods is clear, one cannot readily interpret a particular estimate of noise in economic terms. A monetary value assigned to a noise estimate, on the other hand, would provide clear information about the economic consequences of choice randomness. Second, since the noise parameter is unbounded from above, it is difficult to judge whether the randomness of an agent's

[4] Also known in the literature as the strong utility model or the Fechnerian model.

choices is high or low. A value defined on the unit interval would solve this problem.[5] Third, the raw estimates of μ are not well suited for interpersonal comparisons, since behavioral parameters β also change across people. Having choice randomness expressed in common units, such as money, and taking into account the interpersonal differences in β would help to overcome this issue. These issues arise not only in the standard multinomial logit model but also in its modifications, such as the contextual utility model of Wilcox (2011).

We propose a method of transforming the estimate of μ into a monetary value that addresses these problems by introducing two intuitive measures.[6] The first measure, absolute welfare cost (AWC), puts a dollar value on the choice randomness. It shows how much money, in certainty equivalent terms, an agent would be allowed to "waste" if her choices are rationalized by the underlying structural model. The second measure, relative welfare cost (RWC), is a transformation of the absolute welfare cost that scales it by the monetary value at stake in a choice context. The relative welfare cost is thus defined on the unit interval. It shows what proportion of the total monetary value at stake an agent would be allowed to waste if her choices are rationalized by the model.[7]

Our approach rests on a careful interpretation of the concept of "noise" and "waste." We follow the descriptive, structural literature on risk preferences by assuming a specific model of the manner in which choice randomness is rationalized. In the language of Infante, Lecouteux, and Sugden (2016, p. 21), this is

> ...not an inference about the hypothetical choices of the client's inner rational agent, but rather a way of *regularising* the available data about the client's preferences so that it is compatible with the particular model of decision-making that the professional wants to use. Regularisation in this sense is almost always needed when a theoretical model comes into contact with real data.

[5] The parameter of the tremble model of stochastic choice (Harless and Camerer, 1994) has this property and thus allows one to evaluate the magnitude of the choice randomness. However, its value still would require an economic interpretation.

[6] While the discussion below focuses on the multinomial logit model and its modifications, a similar logic can be applied to other models of stochastic choices, such as the trembles model (Harless and Camerer, 1994) or the random preferences model (Loomes and Sugden, 1995; Gul and Pesendorfer, 2006).

[7] Other ways to measure the welfare costs of stochastic choice might exist, however we find that using monetary measures based on certainty equivalents to be intuitive and transparent.

In our case the subject being evaluated is the "client," and we are the "professional." Thus we consistently use the expression "noise," or some synonym, rather than "error." When it comes to us using this regularised model of the agent, we may then adapt an "intentional stance" towards the evaluation of an agent's behavior, using a philosophical approach developed by Dennett (1987), theoretically interpreted for use in economics by Ross (2014, ch. 4), and explicitly applied to behavioral welfare economics by Harrison and Ross (2018, § 5). Only then can we use the expression "waste." Similarly, when we characterize behavior as being "imperfectly rational" below, that also reflects our intentional stance, rather than a claim that the agent has made an error in cognitive processing or problem representation.

Our absolute and relative welfare cost measures allow one to conveniently evaluate the economic significance of choice randomness, its relative magnitude, and to compare the magnitude of choice randomness across people. The implications of this magnitude for an agent's behavior depend critically on the underlying model of the source of choice randomness.[8] For instance, the Random Utility model (Marschak, 1960) assumes that in each case an agent makes an optimal choice, and the choice randomness is due to the perturbations of the agent's utility function that are unobservable to a researcher. High estimated choice randomness, reflected in the RWC being close to 1, would imply that the stochastic part of the structural model dominates the deterministic part and that the structural model cannot explain the agent's choices well. In this case, the RWC can be viewed as a measure of a model's fit. From the perspective of the Rational Inattention model of Caplin and Dean (2015) and Matêjka and McKay (2015), the source of choice randomness is information costs associated with collecting the payoff-relevant information. In this case, the AWC measure allows one to assign a dollar value to these information costs. Given that the AWC puts choice randomness on the same monetary scale for all subjects, it can further be used to compare people by the magnitude of their information costs and to study the characteristics that predict interpersonal differences in these costs.

[8]This is true even when different "source" models lead to the same choice likelihoods, as for instance is the case for the Random Utility and the Rational Inattention (Caplin and Dean, 2015; Matêjka and McKay, 2015) models that both lead to multinomial logit specifications.

We apply our method to the data from an artefactual field experiment in Denmark. The subjects came from a sample of the general Danish population and were asked to make a series of choices between two risky alternatives. Each subject answered a detailed demographic survey, which we use to characterize the effects of demographic characteristics on the observed heterogeneity in the AWC and RWC. We find that the average AWC is around $8 and thus negligible for the subjects' natural economic environment. However, the RWC is quite significant, and is on average 0.87. There is also considerable variation among the subjects in terms of their AWC and RWC. Regression analysis shows that certain demographic characteristics are associated with higher costs. In particular, subjects who are older, less educated, and have lower income, have larger welfare costs. Females have slightly higher welfare costs than males, but the difference is marginal.

Section 3.2 describes the method of converting an estimate of noise into welfare costs measured in monetary terms and provides an explicit algorithm for computation in a binary choice case. Section 3.3 applies the method to data from an artefactual field experiment in Denmark involving choice under risk and studies the properties of the welfare costs, as well as their demographic correlates. Section 3.4 discusses connections with previous literature. Section 3.5 concludes.

3.2 Method

We first look at an illustrative case, which demonstrates the logic behind our method of extracting the welfare cost information from a noise estimate. Then we turn to a more common binary-choice case and explicitly describe the algorithm to implement this method.

3.2.1 Illustrative Case

Consider an agent choosing from a set of alternatives indexed by real numbers on a compact interval $A = [a_l, a_h]$. Each alternative generates a lottery

$$l(a) = \{x_1(a), \ldots, x_k(a);\ q_1(a), \ldots, q_k(a)\},$$
$$a \in A,\ x_i \in \mathbb{R},\ q_i \in \mathbb{R}_+,\ \forall i = 1, \ldots, k,\ \sum_{i=1}^{k} q_i = 1,$$

where x_i are monetary outcomes and q_i are respective probabilities. To aid interpretation, think of a supplier who is selecting the quality of her product. A higher quality product is more likely to be sold at all the k markets, but with different probabilities for each market. On the other hand, higher quality imposes larger production costs and lower profits in each market.

Each alternative a has an aggregate utility $U(a) \equiv U(l(a))$ associated with it, and outcomes are transformed using $u : \mathbb{R} \mapsto \mathbb{R}$, the von Neumann-Morgenstern utility function. Each value of $U(a)$ can be translated into a certainty equivalent $m(a)$, defined by $u(m(a)) = U(a)$. The u function is assumed to be twice continuously differentiable, strictly increasing and strictly concave, so that its inverse u^{-1} is strictly increasing and strictly convex. The certainty equivalent function $m(a)$ does not preserve concavity, in general, but it is an increasing transformation of $U(a)$, so that the ordering of alternatives is preserved: $U(a) \geqslant U(b) \Leftrightarrow m(a) \geqslant m(b),\ \forall a, b \in A$.

Assume that U is concave and reaches its unique maximum (minimum) at a^* (a_*), as does the certainty equivalent function. Define the maximum certainty equivalent as $m^* \equiv m(a^*)$, and the minimum certainty equivalent as $m_* \equiv m(a_*)$. If the agent always chooses the optimal alternative a^*, we call this behavior *perfectly rational*. On the other extreme, if the agent does not always choose a^*, and moreover, the likelihood of it being chosen is the same as for any other alternative, we call such a behavior *zero rational*. We are concerned with the behavior in between, which is neither perfectly rational nor zero rational, a behavior that we call *imperfectly rational*.

The degree of this imperfection[9] is characterized by a number ε, s.t. $0 \leqslant \varepsilon \leqslant \Delta m$, with Δm being the difference between the maximum certainty equivalent, m^*, and the minimum certainty equivalent, m_*. Choices that lead to certainty equivalents within ε distance from the maximum certainty equivalent can be viewed, from the perspective of a model, as imperfectly rational.[10] These choices form an optimal region A^* defined by

$$A^*(\varepsilon) = \left\{ a \in A \mid m(a) \geqslant m^* - \varepsilon \right\}. \tag{3.2}$$

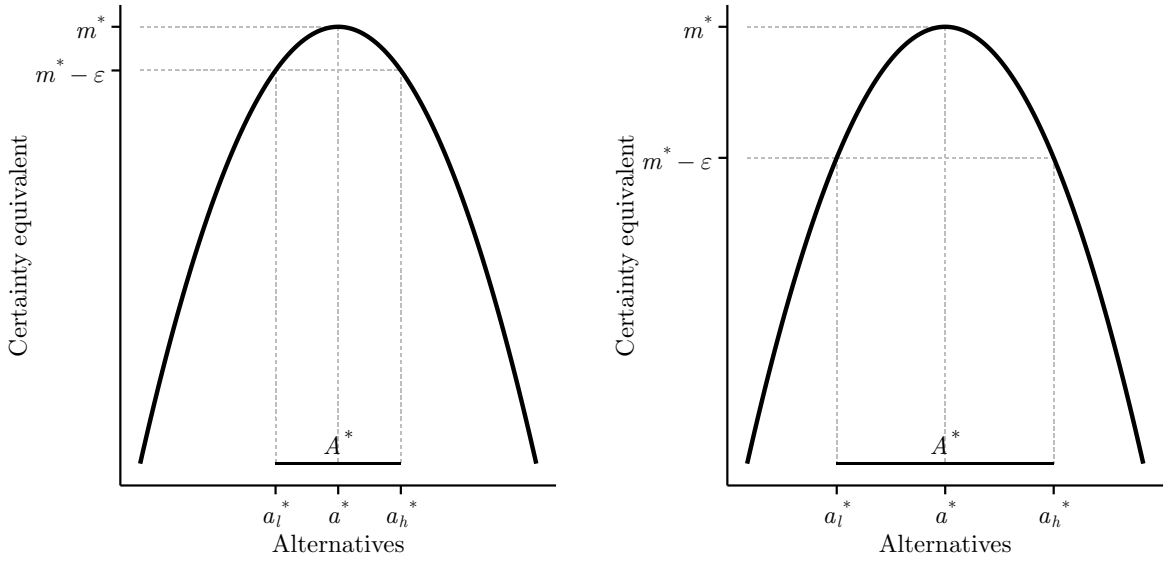
The degree of imperfection, ε, shows how much monetary welfare an agent would be allowed to waste if her choices are rationalizable by the model, and effectively includes these choices in the optimal region. In other words, it represents the welfare costs measured in dollars. Our goal is to link these costs to an estimate of noise.

The higher is the allowed degree of imperfection the wider is the optimal region. If ε is set to 0, the optimal region will consist only of the optimal alternative a^*. If ε is high enough, the optimal region coincides with the whole set of alternatives A. Figure 3.1 illustrates how the optimal region varies with the degree of imperfection. Geometrically, the optimal region is the line segment $\left[a_l^*, a_h^*\right]$.

The optimal region and the degree of imperfection are the first two components that we need to interpret an estimate of noise. The third component comes from the reduced-form stochastic model, which generates a density p over the set of alternatives. Some alternatives fall into the optimal region, by definition. By integrating the density $p(a)$ over this region we get the proportion of choices that are counted, from the perspective of a model, as imperfectly rational for a given

[9] This term should be understood as an imperfection of a given model to regularise data, rather than a statement about an agent making decision errors.

[10] The idea of allowing an agent some degree of imperfection in choices is not new. For example, Harrison (1994) introduces a similar quantity based on an agent's subjective cost of choosing one alternative versus the other to explain many EUT violations.



(a) Low Degree of Imperfection

(b) High Degree of Imperfection

Figure 3.1: Optimal Region and Degree of Imperfection

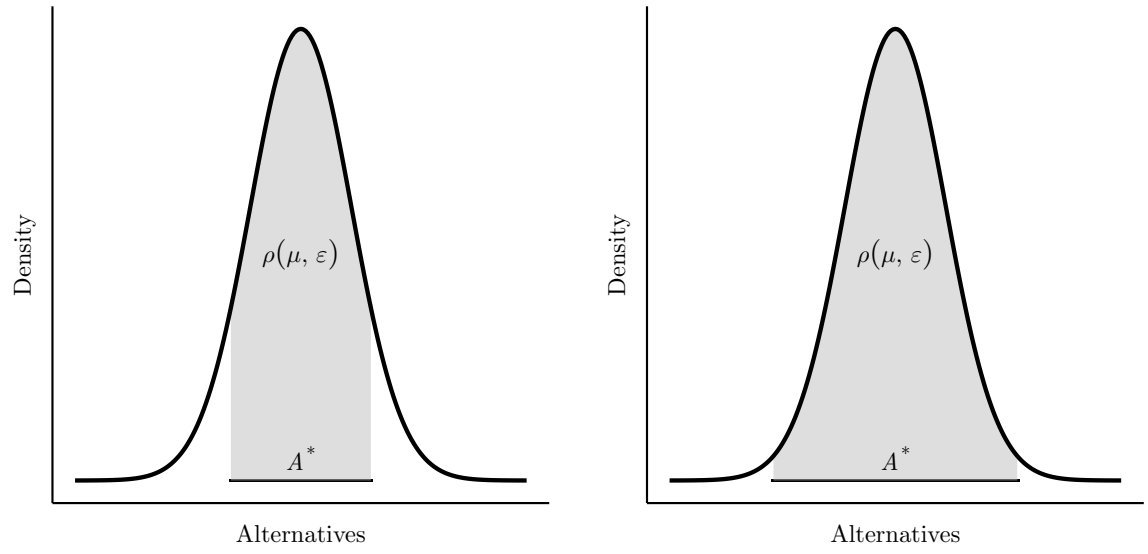
degree of imperfection. We call this measure a *degree of rationality*:[11]

$$\rho(\mu, \varepsilon) = \int_{A^*(\varepsilon)} p(a)\, \mathrm{d}a. \tag{3.3}$$

The degree of rationality has several intuitive properties, two of which turn out to be crucial for our analysis, and can be represented graphically. Figure 3.2 shows that as the degree of imperfection increases, the optimal region expands and the degree of rationality, represented by the gray shaded area, increases. Figure 3.3 shows that as the noise goes up, the density flattens out and the probability mass shifts from the optimal region to the outside area, reducing the degree of rationality.

The degree of rationality for certain values of noise and imperfection has attractive interpretations. The quantity $\rho(\infty, \varepsilon)$ tells us what proportion of choices are counted as rational when they are, in fact, zero rational. It represents a Type II error in a test to detect rationality, and the quantity

[11] It is perhaps more accurate to call it a degree of rationalizability, since we are talking about the choices that can *be rationalized* by the model, given some degree of imperfection ε. Calling it a degree of rationality is somewhat counterintuitive, since it has the property of increasing as imperfection goes up. We retain the shorter term, however, while keeping in mind what this measure actually represents. We will frequently refer to the rationalizability of choices.



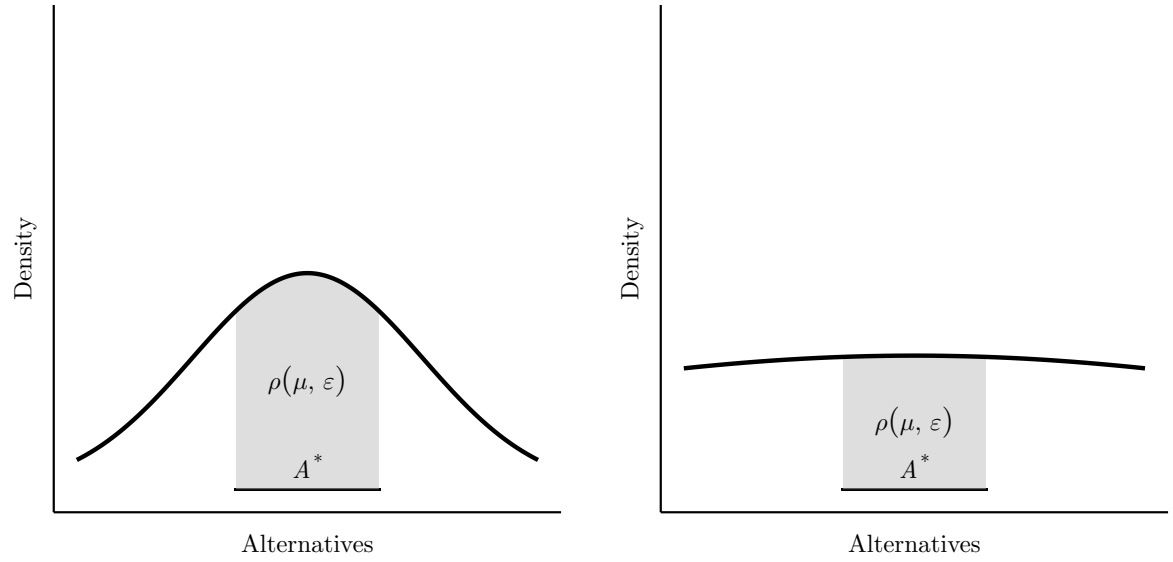
(a) Low Degree of Imperfection (b) High Degree of Imperfection

Figure 3.2: Degree of Rationality and Degree of Imperfection

$1-\rho(\infty,\varepsilon)$ is the power of this test. This power will decrease as the allowed degree of imperfection increases. The degree of rationality $\rho(\hat{\mu},0)$ measures the proportion of rational choices for an estimated level of noise $\hat{\mu}$ and no imperfection. We refer to it as the *default degree of rationality* or default DoR.

We now have all the tools to decipher the noise. We do this by linking a noise estimate, whose value is hard to interpret, to the degree of imperfection, a monetary measure that has an intuitive economic interpretation as the welfare cost, or monetary welfare required to rationalize the agent's choices by a model. In order to link them, we need to reverse the steps we followed so far. Currently, we defined the degree of imperfection, which leads to the optimal region. The optimal region combined with the stochastic model, parameterized by noise, yields the degree of rationality. Suppose that instead we start with the degree of rationality and fix it at some level $\rho = \alpha$. Let an estimated value of the noise be $\hat{\mu}$. The question is how much imperfection should be allowed to for $100 \times \alpha\%$ of the choices to be rationalized for a given noise. In other words, we need to find the ε that satisfies

$$\rho(\hat{\mu},\varepsilon) = \alpha.$$



(a) Low Noise (b) High Noise

Figure 3.3: Degree of Rationality and Noise

This equation establishes an implicit function, $\varepsilon(\hat{\mu};\alpha)$. For the purpose of our analysis, the following property of this function is important.

Proposition 5. *For a given degree of rationality α, the degree of imperfection as a function of noise, $\varepsilon(\mu;\alpha)$, is monotonically increasing:*

$$\frac{d\varepsilon}{d\mu} \geqslant 0.$$

Âě

Proof. See Appendix C.2. □

This property implies that noise and imperfection are in a direct and monotonic relation. It is important since more noise should imply higher welfare costs, which in our case are measured by imperfection. If imperfection and noise were not in a direct and monotonic relation, such an interpretation would be impossible. The relation between ε and μ comes from the fact that the degree of rationality is decreasing in noise and increasing in imperfection. From these properties

it also follows that higher values of α imply higher values of ε. In other words, the more choices we have to rationalize, for a given value of a noise, the more imperfection we should allow.

So far we have treated noise as given and made calculations as though no choices were made, but in practice "noise" is estimated from a series of choices. Consider the case when our supplier makes decisions over the course of several years, indexed by $j = 1, \ldots, n$, and in each year the mapping of quality a into the set of profits and probabilities (lotteries) $l_j(a)$ is different due to changing market conditions. We can repeat all the previous steps in deriving the degree of rationality, but now it will differ by the year: $\rho_j(\mu, \varepsilon)$. Assume that μ does not change over time. We can aggregate the degree of rationality from all the choices by averaging across the degrees of rationality of single choices:

$$\rho(\mu, \varepsilon) = \frac{1}{n} \sum_{j=1}^{n} \rho_j(\mu, \varepsilon). \tag{3.4}$$

Naturally, the average follows all the properties of the degree of rationality for a single choice. In particular, it increases in ε and decreases in μ.

It also makes sense to slightly modify the procedure for computing the value of ε, to take into account the fact that the degree of imperfection should not exceed the difference between the maximum and the minimum certainty equivalents for a given choice, defined as $\Delta m_j \equiv m_j^* - m_{j*}$. Since ε is applied to all the rounds of a decision making process, for some rounds it can actually reach or exceed the Δm threshold, while increasing imperfection beyond this difference does not have any effect on the degree of rationality. This can be addressed by bounding imperfection by the difference in certainty equivalents, and then averaging across all the choices to get an aggregate:

$$\bar{\varepsilon}(\hat{\mu}, \alpha) = \frac{1}{n} \sum_{j=1}^{n} \min\left\{\varepsilon(\hat{\mu}, \alpha), \Delta m_j\right\}. \tag{3.5}$$

We call the resulting measure of imperfection *Absolute Welfare Costs* (generated by noise $\hat{\mu}$, with $100 \times \alpha\%$ of choices rationalized), or AWC. It represents the monetary welfare that the agent would be allowed to give up for exactly $100 \times \alpha\%$ of her choices to be rationalized by the model,

given noise $\hat{\mu}$. For any estimated value of noise and any desired proportion of choices we would like to rationalize we can, therefore, always find a precise dollar value of the welfare costs.

We can go further and translate the welfare costs into relative terms, to compare them with the actual stakes of the choice context. For example, an AWC of \$1 may not look like much, but if the maximum certainty equivalents are around \$1 in all the rounds, it means that almost all the welfare would have to be sacrificed. We divide the degree of imperfection by the difference between the maximum and the minimum certainty equivalents for every round, and average across all the choices:[12]

$$\tilde{\varepsilon}(\hat{\mu},\alpha) = \frac{1}{n}\sum_{j=1}^{n} \min\left\{\frac{\varepsilon(\hat{\mu},\alpha)}{\Delta m_j}, 1\right\}. \tag{3.6}$$

The resulting degree of imperfection represents *Relative Welfare Costs* (generated by noise $\hat{\mu}$, with $100\times\alpha\%$ of choices rationalized), or RWC. Another benefit of this measure is that it allows one to appreciate the magnitude of the noise, since RWC are bound between 0 and 1, while a raw estimate of noise is unbounded from above. If rationalizing $100\times\alpha\%$ of the choices requires on average almost all the difference between the maximum and the minimum certainty equivalents, in which case RWC are close 1, that clearly indicates that the choices are close to being zero rational, from the perspective of the model. On the other hand, if it requires only a small fraction of this difference, in which case RWC are near 0, the choices are close to being perfectly rational, from the perspective of the model.

3.2.2 Binary Choice

An important special case arises when an agent has only two alternatives to choose from. This is one of the most common experimental designs.[13] In this case the set of alternatives in each round is $A = \{a_1, a_2\}$. Without loss of generality, assume that the alternative a_2 always gives the highest utility, so that $U_j(a_2) > U_j(a_1),\ j = 1,\ldots,n$, i.e., $a_j^* = a_2$, using the notational convention

[12] Since the resulting quantity has to be a fraction, we bound this ratio by 1.

[13] The risk-elicitations methods developed and popularized by Holt and Laury (2002) and Hey and Orme (1994) apply to the binary choice case.

$U_j(a) \equiv U\left(l_j(a)\right)$. The maximum and the minimum certainty equivalents are $m_j^* \equiv m_j(a_2)$ and $m_{j*} \equiv m_j(a_1)$, respectively. The optimal region and the degree of rationality can then take only two values:

$$A_j^*(\varepsilon) = \begin{cases} a_2, & \varepsilon < \Delta m_j, \\ A, & \varepsilon \geqslant \Delta m_j, \end{cases} \qquad \rho_j(\mu, \varepsilon) = \begin{cases} p_j(a_2), & \varepsilon < \Delta m_j, \\ 1, & \varepsilon \geqslant \Delta m_j, \end{cases} \tag{3.7}$$

where $p_j(a_2)$ is a choice probability supplied by a stochastic model.

Suppose we observe a subject making a series of binary choices, and estimate a structural model of risk preferences, where $\hat{\gamma}$ is a vector of estimated risk parameters and $\hat{\mu}$ is an estimate of noise. The $\hat{\gamma}$ vector in the Expected Utility Theory (EUT) case is typically just the relative risk aversion, and in the Prospect Theory (PT) case it incorporates the risk aversion parameter(s), the probability weighting parameter(s), and the loss aversion parameter.[14] The computation of AWC and RWC (rationalizing $100 \times \alpha\%$ of the choices) from these data can be performed using the following algorithm.

1. For each round, compute the aggregate utilities of both alternatives, $U_j(a_1; \hat{\gamma})$, $U_j(a_2; \hat{\gamma})$, $j = 1, \ldots, n$.

2. Compute the certainty equivalents of both alternatives, $m_j(a_1), m_j(a_2)$, using the inverse transformation, $m_j(a) = u^{-1}(U_j(a; \hat{\gamma}); \hat{\gamma}), a \in A$, and the difference between them, Δm_j.

3. Compute the likelihoods of choosing each alternative using the stochastic model, $p_j(a; \hat{\gamma}, \hat{\mu}), a \in A$.

4. Start with $\varepsilon = 0$. Compute the degree of rationality in each round using (3.7). Compute the average degree of rationality $\rho(\hat{\mu}, \varepsilon)$ using (3.4).

5. If $\rho(\hat{\mu}, \varepsilon) < \alpha$, increase ε by a small number $\Delta\varepsilon > 0$.

6. Repeat increasing ε until the degree of rationality reaches the target level of α.

[14] The parametrization will depend on the utility and probability weighting functions used. For example, if an expo-power utility function is used, it will have two parameters rather than one.

7. Compute the AWC $\bar{\varepsilon}(\hat{\mu}, \alpha)$ using (3.5). Compute the RWC $\tilde{\varepsilon}(\hat{\mu}, \alpha)$ using (3.6).

3.2.3 Alternative Measures

Note that the proposed computation of welfare costs does not involve actual choices. After estimating risk parameters and noise, we ignore whether the actual choices corresponded to the maximum certainty equivalent or not. A question then arises: what choices do we rationalize, if not actual choices? This question suggests a similar computation, but based on actual choices rather than on likelihoods.

Consider the following alternative algorithm. Start by computing the implied decisions based on the certainty equivalents. Then compare actual and implied decisions by looking at the relative proportion of times when implied and actual decisions coincide. This proportion gives the actual default degree of rationality. Next, calculate the vector of the differences in the certainty equivalents (CE differences) for the cases when implied and actual decisions disagree. These are the "mistakes," from the perspective of the model, we need to "correct," or regularise by adding a structural model of behavioral noise. Start with $\varepsilon = 0$ and increase it by a small positive amount. When ε is lower than the CE difference, the degree of rationality is zero; otherwise it equals one, meaning that implied and actual decisions become equivalent. After that, compute the relative proportion of times when rationalized decisions coincide with the actual ones. Increase ε until this proportion reaches the target level. Compute the average of the bounded (by CE difference) ε for the absolute *actual* welfare costs, and the average of their ratios to CE differences for the relative *actual* welfare costs.

Although this alternative algorithm is almost identical to the previous algorithm, there is a subtle difference. This difference makes us choose in favor of the method described in §3.2.2, which involves rationalizing *potential* choices, as opposed to *actual* ones. Consequently, we obtain the estimates of the potential welfare costs, while the alternative method would give us the actual welfare costs. The key difference between the two methods lies in the fact that the likelihoods of choices represents what could have been chosen if the same options were presented many times.

We view this approach as extracting more information from the same data points. The informational gain is obtained through the introduction of a particular structure that describes the choice probabilities.

Of course, if the two methods gave completely different estimates, one would need a stronger argument in favor of one method against the other. Comparing the potential and actual welfare costs, however, shows that the measures are tightly associated (not reported here). In principle, one could easily substitute one method for another.

Another alternative method of computing the absolute welfare costs would arise if we reconsidered equation (3.5), which involves bounding the value of imperfection by the difference in the certainty equivalents in each round. This is not required and we could, as well, have computed the unbounded absolute welfare costs.[15] One might expect that we would obtain higher estimates of the AWC. Indeed, computations show (not reported here) that the unbounded absolute welfare costs are on average twice as large as the bounded ones, and both measures are tightly associated. We prefer to use the bounded measure, however, since it represents only the welfare costs that can be potentially incurred, while the unbounded measure allows wasting more welfare than there actually is.

3.3 Empirical Analysis

3.3.1 Data

We present the results for 218 adult Danes, a subsample of a larger field study by Harrison, Jessen, Lau, and Ross (2018). The subjects were originally recruited for a survey from two internet-based panels with 165,000 active members combined. Invitations were sent out by email, and the subjects could participate using internet-browsers on their computers or mobile devices. The members of

[15] The computation of the RWC must involve bounding, since they represent a fraction that must lie in the unit interval.

the panels are drawn from the adult population of Denmark between 18 and 75 years of age and come from different regions of the country.

Table 3.1 provides a summary of the socio-demographic characteristics of our subsample, who were invited to participate in experiments after completing the online survey. Slightly less than half of the sample were females and the average age was just less than 50 years. The majority of the sample had college education, and the distribution of income across different income brackets was roughly equal. Most of the participants were either employed as public servants or retired. More than 75% of our sample comes from the Greater Copenhagen area.

The subjects made binary choices across 60 pairs of lotteries and answered a set of demographic questions. Once all the lottery choices were made, one of the choices was selected randomly for payoff. Table C.1 in the Appendix contains the battery of lotteries that were given to the subjects. This battery is based on designs by Loomes and Sugden (1998), Wakker, Erev, and Weber (1994), and Cox and Sadiraj (2008). Together they provide a powerful test of EUT and RDU.

3.3.2 Estimation Procedure

Computation of the welfare costs relies on structural estimates of risk preferences γ and noise μ. We implement the estimation in the standard fashion by maximizing the Bernoulli log-likelihood function at the level of a subject:

$$(\hat{\gamma},\hat{\mu}) = \arg\max_{\gamma,\mu} \sum_{j=1}^{n} \left(y_j \ln p_j(a_2;\gamma,\mu) + (1-y_j) \ln p_j(a_1;\gamma,\mu) \right),$$

where $y_j \equiv \mathbb{I}(a = a_2)_j$ is an indicator variable that takes a value of 1 whenever an alternative a_2 is chosen in round j. The alternative a_2 is taken to be the one on the right side of the screen without loss of generality, and we no longer assume that it gives the highest aggregate utility in all the rounds.

Table 3.1: Socio-Demographic Characteristics of the Sample

Characteristic	Mean
Female	0.46
Age	48.06
Education	
Vocational training	0.19
Low level of formal education	0.21
College, less than 3 years	0.09
College, 3 to 4 years	0.27
College, 5 or more years	0.24
Annual household income, before tax	
Less than 300k DKK	0.23
300k–500k DKK	0.23
500k–800k DKK	0.23
More than 800k DKK	0.17
Not reported	0.14
Occupation	
Public servant	0.42
Student	0.12
Unemployed	0.04
Retired	0.23
Skilled worker	0.03
Unskilled worker	0.06
Self-employed	0.06
Other	0.04
Family	
Has children	0.25
Lives with a partner	0.54
Geographic area	
Copenhagen	0.78
Central Denmark	0.07
Zealand	0.09
Southern Denmark	0.06

We assume that the choice probability $p_j(a_2;\gamma,\mu)$ is given by the strong utility model in the logit form

$$p_j(a_2;\gamma,\mu) = \frac{\exp\left(U_j(a_2;\gamma)/\mu\right)}{\exp\left(U_j(a_2;\gamma)/\mu\right)+\exp\left(U_j(a_1;\gamma)/\mu\right)} = \Lambda\left(\frac{U_j(a_2;\gamma)-U_j(a_1;\gamma)}{\mu}\right),$$

where $\Lambda(\cdot)$ denotes the logistic cumulative density function, and $p_j(a_1;\gamma,\mu) = 1 - p_j(a_2;\gamma,\mu)$.

We also assume that the lotteries are compared according to their expected utilities (dropping an index for the round)

$$U(a;\gamma) = \sum_{i=1}^{k} q_i(a)u(x_i(a);\gamma),$$

and the u function takes the constant relative risk aversion form:

$$u(x;\gamma) = \frac{x^{1-\gamma}}{1-\gamma}.$$

Nothing in our approach relies on assuming an EUT model. In fact, we could have proceeded in a way suggested by Harrison and Ng (2016) and estimated different models for different subjects, classifying our subjects as EUT or RDU, for example. Alternatively, as suggested by Monroe (2017), we could have assumed an RDU model for all the subjects, since correct classification has significant data requirements. Appendix C.1 demonstrates this important generalization by regenerating all results assuming an RDU model of risk preferences, as well as assuming a different utility function, the expo-power utility function, and a different model of stochastic choice, the contextual utility model of Wilcox (2011).

3.3.3 Welfare Costs

Figure 3.4 shows the kernel densities of the estimates of AWC and RWC for three different target levels of degree of rationality (DoR) α. Figure 3.4a shows the densities of AWC in Danish Kroner (DKK).[16] The AWC are strikingly low. For $\alpha = 0.9$, the mean AWC is 42 DKK (approximately

[16] 1 DKK was about 0.15 USD at the time of the study.

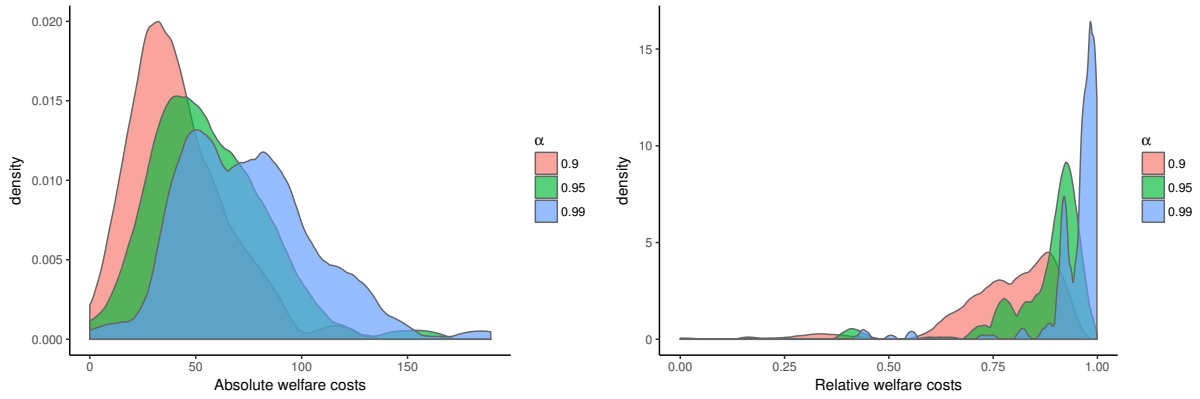
6 USD) and the median is slightly lower at 37 DKK (see Table 3.2). As the target level of DoR increases, the AWC also increase, as expected. The more choices have to be rationalized, the more welfare would have to be sacrificed. This can be seen in Figure 3.4, as the densities corresponding to higher α shift to the right. At the 0.99 level of the target DoR the mean AWC reach 75 DKK (approximately 11 USD), with the median at 71 DKK.

Table 3.2: Summary Statistics for AWC

α	Mean	SD	Min	Q1	Median	Q3	Max
0.9	41.54	22.42	0	26.12	37.35	53.76	123.30
0.95	56.13	26.74	0.96	38.30	52.62	72.10	155.86
0.99	74.80	31.79	5.17	47.78	71.50	92.54	189.25

There is substantial variation among subjects in their AWC. At $\alpha = 0.9$ the smallest AWC is 0 DKK, the absolute minimum one can achieve, while the maximum amount is 123 DKK (approximately 18 USD), which is roughly 3 times as large as the mean value. The standard deviation at this α level is 22 DKK. Several spikes at the right tails of the distributions indicate the existence of subjects with unusually high AWC. The variation in AWC increases as the target level α goes up, which can be seen by the densities becoming flatter in Figure 3.4, as well as by increasing standard deviations and ranges in Table 3.2. At $\alpha = 0.99$ the minimum AWC is still small, just 5 DKK (approximately 1 USD), while the maximum amount becomes 189 DKK (approximately 28 USD), which is again roughly 3 times as large as the mean amount at this level of DoR. The standard deviation reaches 32 DKK. At the highest level of α the distribution of AWC becomes bimodal, which indicates some separation between subjects with relatively low and relatively high AWC.

Figure 3.4b shows densities of RWC. These numbers show that, in contrast to AWC that appear to be tiny for an everyday economic activity, these costs represent a substantial portion of the monetary welfare available in the choice environment. At $\alpha = 0.9$ the mean RWC is 0.78 and the median is 0.81 (see Table 3.3), so more than three quarters of the relative welfare has to be sacrificed in order to rationalize this proportion of choices. Increasing α shifts the distribution of



(a) (b)

Figure 3.4: Absolute and Relative Welfare Costs for 3 Levels of α

RWC to the right, since a larger proportion of rationalized choices requires giving up the whole difference in the certainty equivalents in progressively more choice pairs. At the 0.99 level of α the mean RWC reach 0.95, with the median at 0.97: almost all the welfare must be sacrificed in this case.[17]

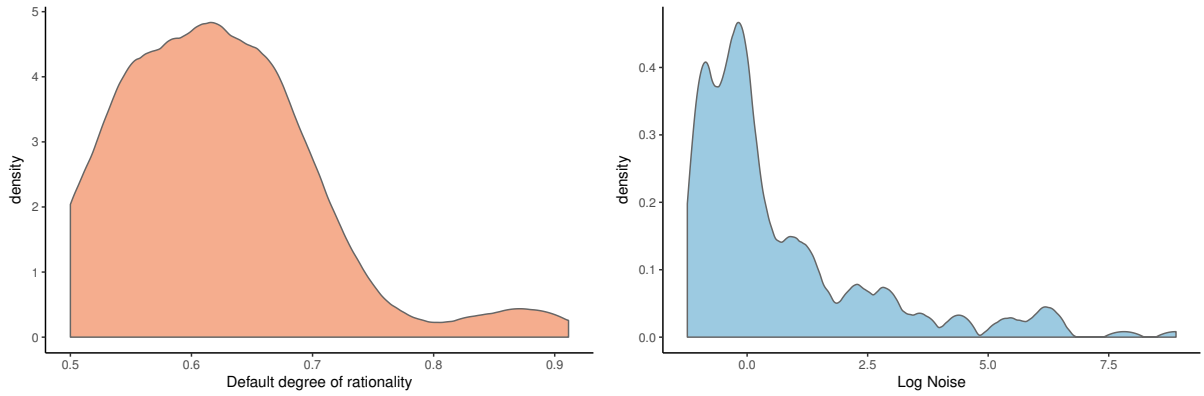
Table 3.3: Summary Statistics for RWC

α	Mean	SD	Min	Q1	Median	Q3	Max
0.9	0.78	0.14	0	0.73	0.81	0.88	0.94
0.95	0.87	0.12	0.16	0.84	0.91	0.93	0.98
0.99	0.95	0.09	0.43	0.93	0.97	0.99	1.00

The RWC numbers also show significant variation across subjects. At $\alpha = 0.9$ the smallest amount of RWC is 0, while the maximum amount is 0.94, which is roughly 1.2 times as large as the mean amount; the standard deviation at this α level is 0.14. There are unusual individual values of the RWC measure, although here these are the subjects who have unusually *low* costs. This can be seen from several distinct spikes on the left tails of the distributions. For some of these outcomes, RWC are less than a half even for the highest level of α. As α increases, the values of RWC become more closely distributed around the maximum of 1, which can be seen by the declining standard deviations in Table 3.3. At $\alpha = 0.99$ the minimum RWC is slightly below

[17]Rationalizing all the choices would definitionally require RWC of 1 for every subject with a non-zero noise, however small, which is the reason to use 0.99 as the highest level of α, and not 1.

a half, 0.43, while the maximum amount becomes 1, and the standard deviation is 0.09. At the highest level of α a similar separation between the subjects with relatively low and relatively high costs arises, represented by two modes in the distribution.



(a) (b)

Figure 3.5: Degree of Rationality and Noise

Figure 3.5a shows the density of the default degree of rationality, $\rho(\hat{\mu}, 0)$. On average, 0.62 of the subjects' choices correspond to the maximum certainty equivalent. This is significantly higher than the minimum of 0.5 (p-value < 0.001 using a t-test), but as the welfare costs estimates show, there is considerable room for improvement. A hump at the right tail of the distribution indicates that there are several extreme cases with higher-than-usual degree of rationality.

The preceding analysis allow us to formulate the following result.

Result 5. *The welfare costs are low in terms of everyday economic activity, but are substantial for the choice environment in which they occurred.*

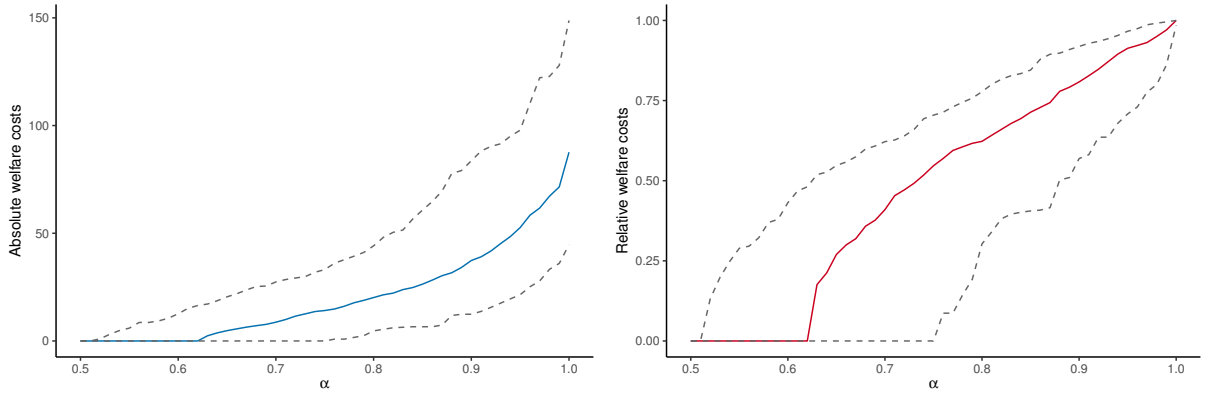
This result suggests that the subjects, on average, were not particularly careful about their choices. This might appear surprising, given that the average expected value of lotteries presented to each subject was 1297 DKK (approximately 195 USD), which is not a trivial income for several hours of work for most people. But the perceived monetary differences between the presented alternatives were not large. The average difference in certainty equivalents between the lotteries was just 94 DKK (approximately 14 USD). On the other hand, there is a substantial heterogeneity in the costs estimates, with some individuals being very careful about their choices.

Comparing our results to Choi *et al.* (2014), we find that the subjects in our sample, in general, made choices of a much lower quality. One possible explanation lies in the differences in methods. The GARP-based measure of choice quality used by Choi *et al.* is well-known for its relatively mild requirements on choice consistency (Beatty and Crawford, 2011). For example, their primary measure does not even require choices to satisfy first-order stochastic dominance.

Another explanation is that there are systematic differences in the samples used. Their sample consisted of adult Dutch individuals, while our sample consisted of adult Danes. We believe this explanation to be unlikely a priori. The results in Blow *et al.* (2008) support our claim, as the study also employs a revealed preference approach and shows that the Danes are generally consistent in other choice domains.

3.3.4 Marginal Welfare Costs

So far we have looked at the distributions of AWC and RWC for only three levels of the target DoR. This analysis does not tell us how quickly welfare costs grow as the rationality requirements become tighter, and in general what the shape of the costs as functions of α is. Figure 3.6 provides an answer to these questions by showing the median welfare costs as functions of α across all the subjects in the sample, with the dashed lines corresponding to the 5% and 95% empirical quantiles.



(a) (b)

Figure 3.6: Absolute and Relative Welfare Costs as Functions of α.

Figure 3.6a shows the graphs of the AWC in relation to α. We observe that the AWC tends to be a convex function of α: at first increasing the DoR requires relatively little AWC, but as the target becomes higher, each additional percentage point of DoR costs more and more in AWC.

The graph for the RWC in Figure 3.6b is in a sense mirror image of the AWC. The RWC tends to be a concave function of α. For small values of DoR extra percentage points of change require high welfare costs, but as the target increases these extra points become less costly in relative terms. These observations allow us to formulate the next result.

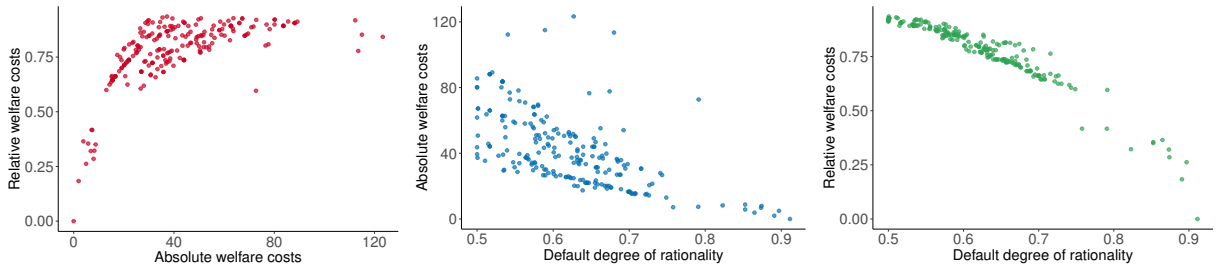
Result 6. *The marginal absolute (relative) welfare costs are increasing (decreasing) with the increase in the target degree of rationality.*

This result can be explained by the way our measures of welfare costs are computed. Starting from a given default DoR, $\rho(\hat{\mu},0)$, we gradually increase ε until the DoR reaches the target, $\rho(\hat{\mu},\varepsilon)=\alpha$. The low marginal AWC at low α targets imply there are quite a few choices that can be easily rationalized by small ε, since the difference in the certainty equivalents between the alternatives must be low. At high target DoR more choices have to be rationalized, but no "easily rationalizable" choices are left. Increasing DoR requires tapping into choice pairs with higher differences in certainty equivalents, and hence higher marginal AWC. The implications for the RWC graphs are the converse. At low target DoR the marginal RWC are high, since rationalizing many choices with small differences in certainty equivalents requires the whole difference. At high targets fewer such choice pairs remain and the marginal RWC decrease.

3.3.5 Relation Between the Measures

We now turn to the relationship between the two welfare costs measures and the default DoR. We ask whether people with lower AWC also have lower RWC, and formally test the previous observation that people with lower DoR tend to have higher costs. The motivation behind these questions is that it is intuitive to expect the positive relation between AWC and RWC. It does not follow, however, directly from the method of their computation. Only if preferences are held

constant must higher AWC imply higher RWC, but there is no such prediction in the case when preferences are not constant, as is typically the case when making comparisons across subjects. Likewise, even though it is natural to expect that people with higher default DoR have lower costs we cannot formally expect this observation to hold *a priori*.



(a) (b) (c)

Figure 3.7: Relation Between the Welfare Costs and Default Degree of Rationality

Figure 3.7a shows a scatterplot of RWC (y-axis) against AWC (x-axis).[18] A clear positive association between the two measures can be observed. The Kendall rank correlation between the two measures is 0.52 and is highly significant, with p-value < 0.001. The relation between RWC and AWC is non-linear, and has a concave shape. The variation in the RWC is highest for intermediate values of the AWC, and gets smaller towards the ends of the range of the AWC.

Figure 3.7b shows the scatterplot of AWC (y-axis) against the default DoR (x-axis), and confirms our earlier observation from the analysis of marginal costs. There is a moderate negative association between the default DoR and the AWC. The Kendall rank correlation between the two measures is -0.52 and is highly significant, with p-value < 0.001. The relation between them is slightly non-linear, and has a convex shape. The variation in the AWC is highest for small values of the default DoR, but declines as the default DoR increases.

Figure 3.7c shows the scatterplot of RWC (y-axis) against default DoR (x-axis). We can immediately see a very tight negative association between the two measures. The Kendall rank correlation between them is -0.84 and is highly significant, with p-value < 0.001. The relation is again slightly non-linear and has a concave shape. The variation in the RWC is almost constant

[18]The analysis in this and next section is conducted for $\alpha = 0.9$. All the reported results hold for $\alpha = 0.95$ and 0.99, though they become less pronounced.

throughout the whole range of the default DoR, and only increases slightly for high values of the default DoR.

These observations allow us to formulate the following result.

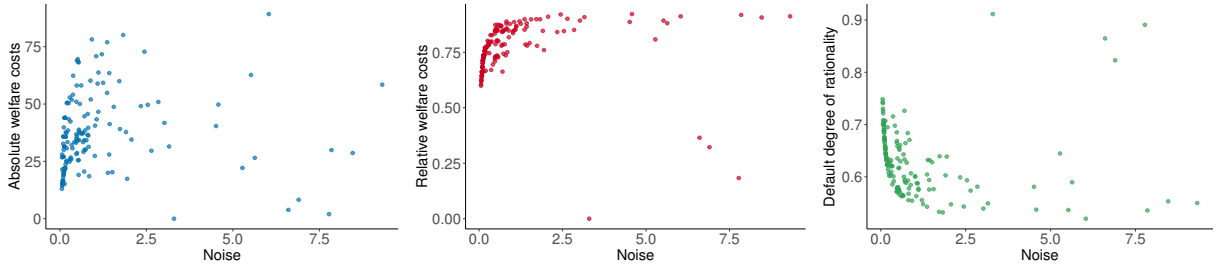
Result 7. *People with higher absolute welfare costs tend to have higher relative welfare costs. People with higher default degree of rationality tend to have lower absolute welfare costs and relative welfare costs.*

This result implies that there is a certain degree of consistency between the measures we introduce. Moreover, this consistency works in the way we expect. This is a nice property, but it could not have been deduced from the method by which these measures are constructed. If risk preferences were the same across subjects, higher AWC must have implied higher RWC, but we cannot say much about the case when preferences and noise are different across subjects. It is possible, that a subject with the high AWC has preferences such that the differences in the certainty equivalents are even higher, and the RWC are actually low. We do, in fact, observe such cases. But the general tendency is for the subjects to have the same ordering, whether it is measured according to the absolute or relative measure of welfare costs.

There is also a negative relation between the default DoR and the welfare two costs measures. This implies that people who make more consistent choices, measured by the default DoR, also require less welfare costs to correct the remaining mistakes. This is an intuitive property, but it is hard to see *a priori* why it should hold even though the data indicate that it does, with the relation between the default DoR and the RWC being particularly strong. The relative strength of this relationship, compared to the relationship with the AWC can be partially attributed to the fact that both the default DoR and the RWC are relative measures defined on the unit interval. Nonetheless, such a strong relationship is remarkable, given that the two measures address two very different questions.

3.3.6 Welfare Costs and Noise

Our approach is in part motivated by the desire to attach an economic meaning to the noise parameter. It is, therefore, of interest to look at the relationship between the two welfare costs measures we introduced and noise, as well as the relationship between the default DoR and noise. Higher noise does translate into higher welfare costs if preferences are kept constant, but no prediction is available for comparisons between subjects, whose preferences are not kept constant. It is natural to expect, however, that this property should also hold between subjects. Given the negative association between the default DoR and the costs, it is also natural to expect that higher noise translates into lower default DoR, but whether it does is an empirical question.



(a) (b) (c)

Figure 3.8: Relation Between the Welfare Costs, the Default Degree of Rationality and Noise

Figure 3.8 shows the scatterplots of (from left to right) the AWC and RWC and the default DoR (on the y-axis) against noise (x-axis).[19] The three panels confirm our hypotheses. We do see that higher noise is associated with higher AWC and RWC and lower default DoR, although the strength of this association differs across the measures. It is small, though statistically significant, for the AWC. The Kendall rank correlation between the two measures is 0.38 (p-value < 0.001).[20] The weakness of the association can be seen by the substantial variation in the AWC at the high values of noise, which means that there are many subjects with high estimates of noise but low AWC. The amount of variation does not allow us to draw conclusions about the likely shape of the relationship between noise and the AWC. The association with the RWC is somewhat stronger,

[19] We restrict noise to be below 10, in order to make Figure 3.8 readable. This excludes around 30% of the subjects. The excluded subjects have noise estimates ranging from around 10 to 7.9×10^8. See Figure 3.5b for the complete distribution of noise estimates.

[20] This and the other two estimates are given for the full sample.

with fewer extreme values. The Kendall rank correlation between the two measures is 0.51 (p-value < 0.001), and follows a concave pattern. Finally, the association with the default DoR is weaker than the association with the RWC, but stronger than the association with the AWC. The Kendall rank correlation between the two measures is -0.44 (p-value < 0.001), and the relation between noise and the default DoR has a convex shape.

A notable feature in these results, most pronounced in the relationship between noise and the default DoR, is that there is an outer boundary that constrains the values. On Figure 3.8c this boundary constrains the values of the default DoR from below and has a convex shape. This means that for given noise the default DoR cannot be lower than a certain value, defined by this boundary.

These findings lead us to the next result.

Result 8. *People with higher noise tend to have higher absolute and relative welfare costs and lower default degree of rationality. For any given value of noise there appears to exist a maximum (minimum) amount of absolute and welfare costs (degree of rationality) that one can have.*

The first part of this result confirms our intuitive guesses, and illustrates further the benefits of the welfare costs measures we propose. The obvious benefit is that they are rooted in economic theory. The data reveals another benefit: our welfare costs measures do not have such an enormous range as noise does,[21] and they allow us to make sensible comparisons between subjects. This benefit becomes clear for subjects with relatively high noise estimates. Despite huge differences in noise between them, their RWC need not be that different. And the variation in the AWC is particularly high, as noted earlier: people might appear to have high welfare costs based on the noise measure, while in fact their AWC are not nearly as large. We do see some association between noise and welfare costs, which implies that noise contains some information about welfare costs and choice consistency, but this information is imprecise.

[21] See Figure 3.5b for the density of the logarithm (base 10) of noise. The logarithmic transformation is used, because for some subjects the estimates of noise are well above 10.

The second part of the result is unexpected and remarkable. It says that there is a regularity in the relation between noise, welfare costs and default DoR. This regularity is in the form of a boundary that constraints the possible values. The existence of such a boundary is related to the estimation and computation procedures, however it is not clear why it exists and what determines its shape. We leave this question for further research.

3.3.7 Who Is More Rational?

We have seen that the estimates of welfare costs and the default DoR vary substantially in our sample. Here we attribute this variability to the observable characteristics of the subjects. We focus on sex, age, education, income, family status, geographic location, and employment status. The demographic covariates are defined as indicator variables, relative to base categories. The base categories are male, age 18–29, vocational training, household income less than 300,000 DKK, no children, living alone, residing in the Copenhagen region, and an unskilled worker.

Table 3.4 provides descriptive regression results. The first three models use the logarithm of AWC as the dependent variable,

$$\ln(AWC)_i = constant + \beta\, Demographic\ controls_i + \varepsilon_i,$$

and are estimated using OLS. Model 1 reports the results for the AWC at $\alpha = 0.9$ and for the basic (no occupation) controls. None of the demographic covariates have significant effects, other than the weakly significant and negative effect of high income. The results suggest, however, that females might have higher AWC than males, and that older and less educated subjects might have higher AWC. Model 2 reports the same results for $\alpha = 0.99$, and shows significant effects of gender. Including occupation controls in Model 3 does not change the results much relative to Model 2, however there is an indication that subjects occupied as public servants have lower AWC on average.

Table 3.4: Regression Results with Demographic Covariates

	Dependent variable:						
	Absolute welfare costs			Relative welfare costs		Default DoR	
	$\alpha = 0.9$	$\alpha = 0.99$	$\alpha = 0.99$	$\alpha = 0.9$	$\alpha = 0.9$		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Constant	3.469***	4.222***	4.459***	1.204***	1.664***	0.557***	0.438**
	(0.202)	(0.162)	(0.259)	(0.254)	(0.450)	(0.111)	(0.173)
Female	0.153	0.195***	0.193**	0.117	0.082	−0.048	−0.031
	(0.093)	(0.075)	(0.076)	(0.116)	(0.116)	(0.051)	(0.052)
Age 30–39	0.084	−0.069	0.014	0.390*	0.579***	−0.177*	−0.251**
	(0.166)	(0.133)	(0.145)	(0.204)	(0.222)	(0.092)	(0.099)
Age 40–49	0.196	−0.040	0.047	0.284	0.481**	−0.136	−0.220**
	(0.168)	(0.135)	(0.155)	(0.205)	(0.234)	(0.093)	(0.106)
Age 50+	0.169	−0.020	0.094	0.396**	0.605***	−0.184**	−0.284***
	(0.149)	(0.120)	(0.152)	(0.181)	(0.226)	(0.083)	(0.104)
Low formal education	−0.050	−0.053	−0.019	−0.224	−0.213	0.098	0.085
	(0.144)	(0.116)	(0.120)	(0.188)	(0.190)	(0.079)	(0.080)
College, $<$ 3 yrs	−0.164	−0.062	0.010	−0.237	−0.113	0.138	0.074
	(0.175)	(0.141)	(0.144)	(0.225)	(0.226)	(0.096)	(0.098)
College, 3–4 yrs	0.009	−0.041	0.030	−0.176	−0.078	0.072	0.024
	(0.134)	(0.108)	(0.112)	(0.177)	(0.178)	(0.073)	(0.075)
College, $>=$ 5 yrs	−0.119	0.024	0.078	−0.465**	−0.396**	0.197**	0.169**
	(0.142)	(0.115)	(0.120)	(0.183)	(0.187)	(0.078)	(0.081)
Income 300k–500k DKK	−0.091	−0.099	−0.047	0.051	0.208	−0.034	−0.102
	(0.134)	(0.108)	(0.121)	(0.171)	(0.189)	(0.073)	(0.082)
Income 500k–800k DKK	0.093	0.066	0.134	0.040	0.239	−0.007	−0.110
	(0.146)	(0.118)	(0.132)	(0.185)	(0.204)	(0.080)	(0.089)
Income $>$ 800k DKK	−0.288*	−0.219	−0.124	−0.421**	−0.159	0.151	0.027
	(0.169)	(0.136)	(0.158)	(0.206)	(0.238)	(0.093)	(0.107)
Income not reported	0.009	−0.013	0.008	0.098	0.161	−0.060	−0.089
	(0.150)	(0.121)	(0.127)	(0.193)	(0.199)	(0.082)	(0.086)
Children	−0.083	−0.026	−0.022	−0.255*	−0.228*	0.129**	0.112*
	(0.112)	(0.090)	(0.092)	(0.138)	(0.137)	(0.062)	(0.063)
Live with partner	0.065	0.029	0.031	0.137	0.132	−0.078	−0.068
	(0.104)	(0.084)	(0.085)	(0.132)	(0.132)	(0.057)	(0.058)
Central Denmark	−0.109	−0.093	−0.112	−0.169	−0.206	0.100	0.104
	(0.180)	(0.145)	(0.149)	(0.224)	(0.225)	(0.100)	(0.101)
Zealand	−0.103	−0.030	−0.015	−0.044	−0.023	0.041	0.022
	(0.155)	(0.125)	(0.126)	(0.197)	(0.194)	(0.085)	(0.085)
Southern Denmark	−0.049	−0.030	−0.086	−0.176	−0.317	0.084	0.142
	(0.195)	(0.157)	(0.159)	(0.247)	(0.249)	(0.107)	(0.108)
Public servant			−0.508**		−1.037**		0.380***
			(0.219)		(0.402)		(0.146)
Student			−0.291		−0.461		0.123
			(0.234)		(0.416)		(0.156)
Unemployed			−0.164		−0.436		0.091
			(0.293)		(0.511)		(0.196)
Retired			−0.444*		−0.749*		0.252*
			(0.229)		(0.421)		(0.152)
Skilled			−0.287		−0.622		0.200
			(0.250)		(0.450)		(0.166)
Self-employed			−0.399		−0.843*		0.403**
			(0.266)		(0.465)		(0.179)
Other occupation			−0.390		−0.556		0.165
			(0.273)		(0.471)		(0.183)
Observations	216	217	217	217	217	217	217

Note: *p<0.1; **p<0.05; ***p<0.01

Models 4 and 5 use the RWC as the dependent variable. Since the RWC are defined only on the unit interval, we a fractional regression model to estimate the coefficients (Papke and Wooldridge, 1996). The models for the RWC are estimated for the basic and extended set of controls and $\alpha = 0.9$.[22] Model 4 shows that older and less educated subjects tend to have higher RWC. Subjects with high income tend to have lower RWC; however, this effect disappears after controlling for occupation in Model 5, which again shows that public servants tend to have lower RWC.

Models 6 and 7 report the results for the default DoR as the dependent variable. It is also defined on the unit interval, and therefore the models are also estimated using a fractional regression model. We report the results for basic and extended controls (recall that the default DoR does not depend on α). The results for default DoR mirror the results for RWC, which is not surprising given the high negative association between the two.

Several patterns emerge from Table 3.4. Age is a significant predictor of RWC and the default DoR. On average, older people have higher welfare costs. Education has a beneficial impact on the welfare costs: in particular, subjects with at least 5 years of college education on average have lower welfare costs and higher default DoR. Subjects with high (more than 800k DKK) income tend to have lower welfare costs. The effect of income disappears, however, if we control for occupation. Subjects with children have (weakly) lower welfare costs and higher default DoR than non-parents. Subjects who are employed as public servants have significantly lower welfare costs and higher default DoR than all other occupations.

We do not see significant differences between males and females in terms of RWC or default DoR. However, using AWC as the dependent variable and $\alpha = 0.99$ (Models 2 and 3), we find that females have larger welfare costs. The geographic region in which subjects live and civil status (whether a subject lives with a partner or not) do not explain much of the variation in welfare costs.

These observations lead us to the following result.

[22] For the AWC and RWC the results for other α's are qualitatively similar, though slightly less pronounced.

Result 9. *Having higher welfare costs and lower default degree of rationality is associated with higher age, lower education, and lower income. RWC are not significantly different for males and females, although the AWC for females tends to be higher.*

This result suggests that formal education is important for the consistency of choices. People with more years of education make more careful choices and, therefore, incur smaller welfare costs. One possible explanation of this result is that exposure to college may produce better analytical skills. On the other hand, if going to college represents a signaling story, i.e., people with higher ability pay for more years of college to signal their ability, this result implies that our measures of welfare costs are correlated with ability. Support for the latter explanation is provided by the effect of income. We find that subjects with higher incomes tend to have lower welfare costs. Therefore, if subjects with high ability are compensated with a higher income, which is reasonable to assume as a first approximation, this result implies that our welfare costs measures are, in fact, correlated with ability. This hypothesis is also consistent with the results for age, since certain mental abilities tend to decline with age after reaching a peak between 20 and 30 years old (Deary and Der, 2005).

3.4 Related literature

Our approach connects to a large theoretical literature on stochastic choice, which we briefly summarize. The early work on stochastic choice dates back to Fechner (1860) and Thurstone (1927). It was subsequently developed into the Random Utility Model (RUM) by Marschak (1960) and summarized by McFadden (2001). Luce (1959) introduced and axiomatized the strong utility (or multinomial logit) model, as well as other models of stochastic choice. McFadden (1976) established necessary and sufficient conditions under which a RUM is equivalent to the multinomial logit model.

Wilcox (2011) extends the standard multinomial logit model by allowing for the noise heterogeneity that is caused by the range of monetary stakes in a choice context. This extension allows one to preserve the deterministic notion of being more risk averse in a stochastic setting. The

stronger utility model developed by Blavatskyy (2014) also allows for noise heterogeneity, but focuses on preserving the first-order stochastic dominance relation in a stochastic choice setting. Gul, Natenzon, and Pesendorfer (2014) modify the multinomial logit model by considering the attributes of choice alternatives rather than alternatives themselves, to address some of the criticism of the original formulation. Apesteguia, Ballester, and Lu (2017) characterize the RUM that satisfies a single-crossing property.

Conceptually, our measures are similar to the Critical Cost Efficiency Index (CCEI) of Afriat (1972), which is used to evaluate the degree of consistency with the Generalized Axiom of Revealed Preference (GARP). Just like our relative cost measure, CCEI is defined on the unit interval, and its complement shows what proportion of monetary value an agent should be allowed to waste in order to rationalize her choices by some utility function. While GARP provides qualitative statements, we put more structure on it in a flexible manner to complement it and provide quantitative evidence.

Viewing our approach as a structural extension of GARP allows us to position our approach again in a broader methodological setting. Ross (2014, ch. 4) carefully lays out the full case for interpreting economic experimentation as an application of the intentional stance of Dennett (1987), noted earlier. This is the methodology that Ross (2014) calls "neo-Samuelsonian," a label that tries to nudge economists toward seeing that the intentional stance is what they have always been doing when they applied Revealed Preference Theory to actual, finite, choice data. In other words: our approach is not novel, exotic economic methodology. Instead we view it as just a sophisticated, structural interpretation of the good old-time religion for economists.

The intuition behind the computation of our measures also links it to a literature on payoff dominance in experiments (Harrison, 1994, 1992; Harrison and Morgan, 1990; Harrison, 1989). This literature shows that by allowing for small deviations from optimal behavior, just as we do, allows one to rationalize supposedly anomalous effects observed in experimental studies.

Harrison and Ng (2016) use an approach similar to ours in order to evaluate the loss of consumer surplus resulting from suboptimal insurance choices. Harrison and Ross (2018) apply the

same approach to evaluate suboptimal portfolio investments. Their measure of lost consumer surplus is similar to our AWC measure, with both being based on computing certainty equivalents. One important difference, however, is that these studies use two experimental tasks: one for preference estimation and the other for welfare evaluation, while we rely on a single task to estimate welfare costs resulting from stochastic choice. The approach that we take in this study does not rely on an independent risk metric, as is the case in Harrison and Ng (2016) and Harrison and Ross (2018), but rather relies on a specific noise structure to "bootstrap" a measure of welfare costs.

Our approach is closely related to studies that estimate structural models of choice under risk and over time. Holt and Laury (2002) study subjects' choices under risk in a laboratory experiment. Subjects make choices between a "safe" and a "risky" lottery across different pairs of lotteries, in which the probabilities of lottery outcomes vary from one pair to the next. HL estimate the Expected Utility model with a flexible Expo-Power utility function using the strict utility model of stochastic choice.[23] Andersen, Harrison, Lau, and Rutström (2008) also use the strict utility model to structurally estimate risk and time preferences of a representative sample of the Danish population. They note that noise estimates are higher in the risk task than in the discounting task. von Gaudecker, van Soest, and Wengstrom (2011) uses a representative sample of the Dutch population to estimate subjects' risk preferences using a model of stochastic choice that is a hybrid between the multinomial logit and tremble models, and thus features two measures of choice randomness: noise and trembles. While these studies typically focus on estimates of risk and time preferences, and do not interpret the estimates of the stochastic part, such as noise or tremble parameters,[24] we explicitly focus on the estimates of the stochastic part and provide a systematic approach to economically interpret the estimates of choice randomness. Finally, Bland (2018) considers mixture specification over pooled choices, contrasting one "rational" model as one of the data generating processes (DGP) with a "behavioral" model as the other DGP. He then calculates CE of choices us-

[23] The strict utility model (Luce, 1959) differs from the multinomial logit model in the way the noise parameter enters choice likelihoods.

[24] von Gaudecker, van Soest, and Wengstrom (2011) is an exception, which provides a brief discussion of the economic significance of the estimates of the tremble parameter. In particular, they give an example of what the estimated parameters of the stochastic part of the model imply for the relation between the difference in the certainty equivalents and the likelihood of choosing the higher-valued lottery.

ing the deterministic core of the "rational" model DGP, thereby evaluating potential welfare losses from using a "behavioral" DGP as well as the existence of noise for both DGP. We reject the simplistic identification of one model as "rational" and the explicit assumption that the "behavioral" model is therefore "irrational." But the general logic of allowing the estimated structural model of noise to provide a basis for welfare evaluation is consistent with our approach.

We provide economic measures of choice randomness (or consistency), which link to studies on the quality of decision-making. Choi, Fisman, Gale, and Kariv (2007) study decision-making under risk in a laboratory experiment in which they present subjects with convex budget sets for two Arrow securities. This design allows them to gauge the subjects' decision-making quality using a measure of GARP-consistency, a standard technique in the revealed preference approach to consumer demand. They find that subjects' behavior is highly consistent with GARP. Choi, Kariv, Müller, and Silverman (2014) expand the analysis by using a representative panel of the Dutch population. They also find a high degree of GARP-consistency in risky choices, which varies, however, with education, sex, and age. Beatty and Crawford (2011) show that while behavior in a wide range of situations is highly GARP-consistent, this might be a result of a misspecified measure of consistency. They propose an alternative to the traditional CCEI measure, which is based on predictive success, and show that the CCEI measures of GARP-consistency are overinflated, and hence that the actual consistency of choices is much lower. Hey (2001) studies decision-making quality in a laboratory experiment on choice under risk and asks whether choice consistency improves with experience. He finds mixed evidence of a positive effect of experience on choice consistency. We rely on a parametric measure of choice consistency and find a lower degree of consistency than in the studies that use the non-parametric revealed preference approach.

Finally, our approach is also related to recent literature on rational inattention. Matêjka and McKay (2015) show that when an agent faces information costs, optimal behavior is stochastic choice, and that under certain conditions choice likelihoods are represented by the multinomial logit specification. Cheremukhin, Popova, and Tutino (2015) apply a model of rational inattention to risky choices and estimate the shape of the cost-of-information function in a laboratory experi-

ment with student subjects. Caplin and Dean (2015) develop a revealed preference test of rational inattention theories with general cost-of-information functions. Since the noise parameter in the rational inattention models has the interpretation of marginal information costs, our method allows one to convert these costs into monetary or percentage terms.

3.5 Conclusion

Stochastic choice has become an active area of both theoretical and empirical research. While the existing literature mainly focuses on the sources of choice randomness, its economic consequences are less well understood. We develop tools to assess the economic significance of noise and apply them to a sample from the general Danish population in an artefactual field experiment.

We introduce three interconnected concepts: rationalizing imperfection, optimal region, and degree of rationality. Fixing the degree of rationality at a certain target level, we vary the amount of imperfection, which in turn affects the optimal region, to make the proportion of subjects' choices falling in the optimal region equal the target level. This amount of imperfection represents the welfare costs, or monetary welfare allowed to be wasted, that is required to rationalize by a model a given proportion of choices. The resulting welfare costs can be expressed both in absolute (dollar) and in relative (to the actual stakes of the choice environment) terms.

We compute the absolute welfare costs and relative welfare costs at the individual level in an experiment with binary-choice lotteries. Several patterns emerge from our analysis, some of which coincide with previous findings, and some of which are new. We find that the AWC are not economically significant in our sample, while the RWC are economically significant. In other words, the welfare costs are tiny if viewed from a broad perspective of economic activity, but they are substantial if viewed from the perspective of this particular choice experiment. As compared to Choi *et al.* (2014), who employ a relative measure based on the consistency with GARP, our estimates of RWC are much larger. However, our results for choice consistency are comparable with that of von

Gaudecker *et al.* (2011), who also employ structural methods. We attribute the difference in results to the difference in the methods, with our method imposing stricter requirements on rationality.

Since our welfare costs measures depend on the target level of rationality α, we study the shape of the relation between α and these welfare costs. We find that the AWC increase in α at an increasing speed, while the RWC increase in α at a decreasing speed. The difference in these two relations is explained by the way our method of computation works. Regardless of the α, however, we observe that subjects' welfare costs functions are roughly "parallel": for any given level of the target degree of rationality the subjects tend, on average, to preserve their ordering by the welfare costs. This can be shown by the relation between the welfare costs measures and the default degree of rationality. Subjects with higher AWC tend to have higher RWC. Also, a lower default degree of rationality is associated with higher AWC and RWC: subjects who start out with low default degree of rationality require a higher cost to reach a given degree of rationality. Looking at the relation between our cost measures and raw estimates of noise reveals that they are positively associated, though our measures do not have such a wide range, which allows for sensible comparisons across subjects and allows us to make judgements about the magnitudes of choice inconsistencies.

The analysis of observable heterogeneity and its role in predicting welfare costs suggests patterns similar to those reported by von Gaudecker *et al.* (2011) and Choi *et al.* (2014). We find that welfare costs increase with age, decline with education, and decline with income. This pattern of results suggests that our measures of costs may be correlated with subjects' inherent abilities.

Finally, we take seriously the need for consistent methodological and philosophical positions when it comes to undertaking behavioral welfare economics. The reason is simple: one cannot question the consistency of observed choices by agents on the one hand and then turn around and effortlessly infer the preferences of those agents on the other hand. This isolates the deep normative challenge raised by the core descriptive insight of behavioral economics, as stressed by Ross (2014, ch. 4), Infante, Lecouteux, and Sugden (2016), and Harrison and Ross (2018, § 5). Dennett (1987)'s intentional stance, as applied to economics by Ross (2014)'s "neo-Samuelsonian" methodology, provides a general and consistent approach to address this challenge, and permits

concrete applications illustrated by Harrison and Ng (2016), Harrison and Ross (2018) and the present study.

Appendix A

Chapter 1

A.1 Treatments Used in the Effort Task

Table A.1: Summary of Treatments

id	θ	w	z	k	5_21_15	5_28_15	6_2_15	6_3_15	6_4_15	6_5_15
1	0	1	2	1	1	1	1	1	1	1
2	0	1	4	1	1	1	1	1	1	1
3	0	2	2	1	1	1	1	1	1	1
4	0	2	2	2	1	1	1	1	1	1
5	0	2	4	1	1	1	1	1	1	1
6	0	2	4	2	1	1	1	1	1	1
7	0.25	1	2	1	1	1	1	0	0	0
8	0.25	2	2	1	0	0	0	1	1	1
9	0.25	2	2	2	0	0	1	0	0	0
10	0.25	2	4	1	0	0	0	1	1	1
11	0.50	1	2	1	1	1	1	0	0	0
12	0.50	2	2	1	0	0	1	1	1	1
13	0.50	2	2	2	0	0	1	0	0	0
14	0.50	2	4	1	0	0	0	1	1	1
15	0.75	1	2	1	1	1	0	0	0	0
16	0.75	2	2	1	0	0	1	1	1	1
17	0.75	2	2	2	0	0	1	0	0	0
18	0.75	2	4	1	0	0	0	1	1	1
19	1	1	2	1	1	1	1	1	1	1
20	1	1	4	1	1	1	1	1	1	1
21	1	2	2	1	1	1	1	1	1	1
22	1	2	2	2	1	1	1	1	1	1
23	1	2	4	1	1	1	1	1	1	1
24	1	2	4	2	1	1	1	1	1	1

Notes: The first column is an id of a treatment, and the next four columns show the values of the treatment variables for that treatment. The last six columns correspond to the six sessions and indicate whether a treatment was (1) or was not (0) used in a session.

A.2 The Battery of Lotteries Used in the Lottery Task

Table A.2: The Battery of Lotteries

ID	Lp1	La1	Lp2	La2	Lp3	La3	Rp1	Ra1	Rp2	Ra2	Rp3	Ra3
1	0	0	1	3	0	13	0.01	0	0.89	3	0.10	13
2	0.60	0	0	18	0.40	35	0	0	1	18	0	35
3	0.80	8	0	15	0.20	23	0.70	8	0.30	15	0	23
4	0.60	0	0	18	0.40	35	0.15	0	0.75	18	0.10	35
5	0.90	0	0	18	0.10	35	0.75	0	0.25	18	0	35
6	0.10	0	0	18	0.90	35	0	0	0.20	18	0.80	35
7	0.50	0	0	18	0.50	35	0.10	0	0.80	18	0.10	35
8	0.50	0	0	18	0.50	35	0	0	1	18	0	35
9	0.10	0	0.80	18	0.10	35	0	0	1	18	0	35
10	0.70	0	0	18	0.30	35	0.50	0	0.40	18	0.10	35
11	0.70	0	0	18	0.30	35	0.40	0	0.60	18	0	35
12	0.50	0	0.40	18	0.10	35	0.40	0	0.60	18	0	35
13	0.89	0	0.11	3	0	13	0.90	0	0	3	0.10	13
14	0.90	0	0	18	0.10	35	0.80	0	0.20	18	0	35
15	0.10	0	0	18	0.90	35	0	0	0.25	18	0.75	35
16	0.40	0	0	18	0.60	35	0.10	0	0.75	18	0.15	35
17	0.40	0	0	18	0.60	35	0	0	1	18	0	35
18	0.10	0	0.75	18	0.15	35	0	0	1	18	0	35
19	0.70	0	0	18	0.30	35	0.60	0	0.25	18	0.15	35
20	0.70	0	0	18	0.30	35	0.50	0	0.50	18	0	35
21	0.60	0	0.25	18	0.15	35	0.50	0	0.50	18	0	35
22	0.85	0	0	18	0.15	35	0.75	0	0.25	18	0	35
23	0.10	0	0	18	0.90	35	0	0	0.30	18	0.70	35
24	0	0	1	8	0	38	0.01	0	0.89	8	0.10	38
25	0.40	0	0	18	0.60	35	0.20	0	0.60	18	0.20	35
26	0.40	0	0	18	0.60	35	0.10	0	0.90	18	0	35
27	0.20	0	0.60	18	0.20	35	0.10	0	0.90	18	0	35
28	0.60	0	0	18	0.40	35	0.50	0	0.30	18	0.20	35
29	0.60	0	0	18	0.40	35	0.40	0	0.60	18	0	35
30	0.50	0	0.30	18	0.20	35	0.40	0	0.60	18	0	35
31	0.80	0	0	18	0.20	35	0.70	0	0.30	18	0	35
32	0.10	0	0	18	0.90	35	0	0	0.40	18	0.60	35
33	0.25	0	0	18	0.75	35	0.10	0	0.60	18	0.30	35
34	0.25	0	0	18	0.75	35	0	0	1	18	0	35
35	0.89	0	0.11	8	0	38	0.90	0	0	8	0.10	38
36	0.10	0	0.60	18	0.30	35	0	0	1	18	0	35
37	0.50	0	0.20	18	0.30	35	0.40	0	0.60	18	0	35
38	0.55	0	0	18	0.45	35	0.40	0	0.60	18	0	35
39	0.55	0	0	18	0.45	35	0.50	0	0.20	18	0.30	35
40	0.70	0	0	18	0.30	35	0.60	0	0.40	18	0	35
41	0.15	0	0	18	0.85	35	0	0	0.25	18	0.75	35
42	0.30	0	0	18	0.70	35	0.15	0	0.25	18	0.60	35
43	0.30	0	0	18	0.70	35	0	0	0.50	18	0.50	35
44	0.15	5	0	10	0.85	30	0	5	0.25	10	0.75	30
45	0.30	5	0	10	0.70	30	0.15	5	0.25	10	0.60	30
46	0.30	5	0	10	0.70	30	0	5	0.50	10	0.50	30
47	0.15	0	0.25	18	0.60	35	0	0	0.50	18	0.50	35
48	0.15	5	0.25	10	0.60	30	0	5	0.50	10	0.50	30
49	0.15	5	0.75	10	0.10	30	0	5	1	10	0	30
50	0.60	5	0	10	0.40	30	0	5	1	10	0	30
51	0.60	5	0	10	0.40	30	0.15	5	0.75	10	0.10	30
52	0.90	5	0	10	0.10	30	0.75	5	0.25	10	0	30
53	0.10	3	0	13	0.90	28	0	3	0.25	13	0.75	28
54	0.40	3	0	13	0.60	28	0.10	3	0.75	13	0.15	28
55	0.40	3	0	13	0.60	28	0	3	1	13	0	28
56	0.10	3	0.75	13	0.15	28	0	3	1	13	0	28
57	0.70	3	0	13	0.30	28	0.60	3	0.25	13	0.15	28
58	0.15	0	0.75	18	0.10	35	0	0	1	18	0	35
59	0.70	3	0	13	0.30	28	0.50	3	0.50	13	0	28
60	0.60	3	0.25	13	0.15	28	0.50	3	0.50	13	0	28
61	0.85	3	0	13	0.15	28	0.75	3	0.25	13	0	28
62	0.10	8	0	15	0.90	23	0	8	0.30	15	0.70	23
63	0.40	8	0	15	0.60	23	0.20	8	0.60	15	0.20	23
64	0.40	8	0	15	0.60	23	0.10	8	0.90	15	0	23
65	0.20	8	0.60	15	0.20	23	0.10	8	0.90	15	0	23
66	0.60	8	0	15	0.40	23	0.50	8	0.30	15	0.20	23
67	0.60	8	0	15	0.40	23	0.40	8	0.60	15	0	23
68	0.50	8	0.30	15	0.20	23	0.40	8	0.60	15	0	23

Notes. The columns are coded as follows: “L" and “R" denote left and right lottery, “a" denotes amounts (in $) and “p" denotes probabilities.

A.3 Math Quiz

1. Suppose you throw a 10-sided die once. What is the probability that any particular number $(1, 2, \ldots, 10)$ will come up?

   - 1%
   - 5%
   - 10%
   - 25%
   - 50%

2. Suppose you throw a 20-sided die once. What is the probability that any of the numbers $1, 2, 3, 4, 5$ will come up?

   - 1%
   - 5%
   - 10%
   - 25%
   - 50%

3. Suppose you throw two 10-sided dice once. One die has numbers $0, 10, 20, \ldots, 90$ on it and the other die has numbers $0, 1, 2, \ldots, 9$ on it. What is the probability that the sum of the numbers on the two dice will be exactly 71?

   - 1%
   - 5%
   - 10%
   - 25%
   - 50%

4. Suppose you flip a fair coin 100 times. You get 2 pennies if tails come up and you lose a penny if heads come up. What would be the expected result of this game for you?

   - Lose 50 pennies
   - Lose 20 pennies
   - Neither gain nor lose anything
   - Gain 20 pennies
   - Gain 50 pennies

5. Let's say there are two games of flipping coins to play. In the first game you get 10 dollars if tails come up and lose 5 dollars if heads come up. In the second game you get 11 dollars if tails come up and lose 4 dollars if heads come up. Which of these games would be more beneficial for you?

- First game
- Second game

6. Let's say you have two games to play. In the first game, you flip a coin and get a dollar if it's tails and lose a dollar if it's heads. In the second game, you throw a 10-sided die and get a dollar if any of the numbers $1, 2, 3, 4$ come up and lose a dollar if any other number comes up. Which game would be more beneficial for you?

    - First game
    - Second game

A.4 Demographic Survey

1. What is your age?
2. What is your gender?

    - Male
    - Female

3. What is your racial or ethnic background?

    - White or caucasian
    - Black or African American
    - Hispanic
    - Asian
    - Native American
    - Multiracial
    - Other
    - Prefer not to answer

4. What is your marital status?

    - Married
    - Single
    - Divorced
    - Widowed
    - Other
    - Prefer not to answer

5. What is your major/field of study?

- Accounting
- Economics
- Finance
- Business Administration
- Education
- Engineering
- Health and Medicine
- Biological and Biomedical Sciences
- Math, Computer Sciences, or Physical Sciences
- Social Sciences or History
- Law
- Psychology
- Modern Languages and Cultures
- Other

6. What is your GPA?
7. What is your year in school?
    - Freshman
    - Sophomore
    - Junior
    - Senior
    - Masters
    - Doctoral
8. What is the number of people in your household?
9. What is the total income of your household?
    - Under $5000
    - $5000—$15000
    - $15001—$30000
    - $30001—$45000
    - $45001—$60000
    - $60001—$75000
    - $75001—$90000
    - $90001—$100000

- Over $100001
- Prefer not to answer

10. What is the total income of your parents?

    - Under $5000
    - $5000—$15000
    - $15001—$30000
    - $30001—$45000
    - $45001—$60000
    - $60001—$75000
    - $75001—$90000
    - $90001—$100000
    - Over $100001
    - Don't know
    - Prefer not to answer

A.5 Subject Instructions

Introduction

Welcome and thank you for participating! This is an experiment in individual economic decision-making. Please, mute/turn off all of your electronic devices for the duration of the experiment.

Payment

Your total payment will consist of a participation payment of $5 and the sum of the payments from the two decision tasks. Your payment can be considerable and will depend on your decisions and chance. You will be paid in cash privately at the end of the session.

Time

Today's session will consist of a quiz on probabilities, two decision tasks and a demographic survey. The session will take up no more than 2 hours.

Payment Protocol

Each of the 2 decision tasks will have several rounds. At the end of each task we will randomly select one of the decision rounds to determine your payment. It is worthwhile to think carefully about each decision, since you don't know which decision round will be picked.

Privacy

You will not interact with other participants. Please, do not reveal your identity to anyone. You must not talk to other participants during the experiment.

Final Notes

Please read these instructions carefully. You are welcome to ask questions at any point. Just raise your hand and we will answer your question in private.

Task 1

In this task you will choose an effort level for a project (Figure 1). A project has two possible outcomes: success or failure. In case of success the project will yield you a high revenue (i.e., payoff), in case of failure it will yield you a low revenue. The exact values of high and low revenues will be shown on the screen. The task is to select the level of effort you prefer the most. There are no right or wrong answers, just pick whatever suits you the most.

Effort

By choosing a higher effort level you increase the chances that the project will be successful. Equivalently it means that the chances of failure are reduced, because the probability of success and failure must add up to 100%. Effort is costly to you: the higher is the effort level, the higher is the cost. The cost will be subtracted from the revenue of the project. On the screen you can observe how chances of success/failure, cost of effort and the profit (=revenue minus cost) change as you change the effort level.

Example

On the Figure 1 you can see a project that gives you a revenue of $30 if it's successful and $10 if it fails. Suppose you chose an effort level of 40%, which leads to a 55% probability of success (45% probability of failure) and costs you $1.60. In case of success your profit will be: $30 (revenue) - $1.60 (cost of effort) = $28.40 (profit). If the project fails your profit will be: $10 (revenue) - $1.60 (cost of effort) = $8.40 (profit). Note that you bear the cost of effort regardless of whether the project succeeds or fails.

Each additional unit of effort will increase the probability of success by the same amount but will cost you more than the previous one.

Example

Increasing effort from 0% to 1%, or from 1% to 2%, or from 2% to 3% (and so on) increases the probability of success by the same amount. Increasing effort from 99% to 100% costs more than an increase from 98% to 99%, which in turn costs more than an increase from 97% to 98% (and so on).

Difficulty

Another important characteristic of the project is its difficulty. Difficulty affects the chances of success, just like effort, but in the opposite way. A more difficult project is less likely to succeed than an easier one, for any given level of your effort. The difficulty of the project will appear on the top of the choice screen.

Example

In the previous example, suppose the difficulty was 30%. Now imagine that the project's difficulty increased to 70%. Given the same effort level as before, 40%, the probability of success might decrease to 25% (equivalently, the probability of failure might increase to 75%). Note that difficulty does not affect revenues or cost of effort, only the chances of success/failure.

Payoff

The outcome of the project will be determined right after you submit your choice. You will see a bar with a success region, a failure region and a randomly moving white needle (Figure 2). The regions are determined by your choice of effort. The needle is equally likely to appear at any position along the bar. It will stop after 3 seconds. If it ends up in the success region the project will succeed, if it ends up in the failure region the project will fail (Figure 3). The next screen will show you the summary of the current round (Figure 4).

Example

On the Figure 2 the probability of success is 55% and the probability of failure is 45%. If at the moment you stop the needle it is in the region 0–55, the project will succeed (Figure 3). If it is in the region 55.1–100, it will fail.

There will be 15 rounds, as well as 5 practice rounds. Rounds will differ in the project difficulty, costs and possible revenues. After you complete all the rounds, one of them will be randomly picked for payoff. You will see bars with rounds' numbers and their results (Figure 5). These bars will be randomly highlighted. Each bar is equally likely to be highlighted at any given moment. It will stop after 3 seconds and the payoff will be determined.

Task 2

In this task you will choose between the two lotteries: left and right (Figure 6). Each lottery offers monetary prizes with some probabilities. You will be asked to choose a lottery you prefer the most. There are no right or wrong answers in this task, the choice of a particular lottery is a matter of personal taste.

Example

Look at the Figure 6, it shows two lotteries. The left lottery offers $8 with 100% chance. The right lottery offers $0 with 1% probability, $8 with 89% probability and $38 with 10% probability.

This task will consist of several rounds. The exact number of rounds will be shown on the top of the screen. Rounds will differ by the lotteries offered, they will have different monetary prizes and probabilities of getting these prizes.

Payoff

Once you complete all the rounds one of them will be randomly picked for payoff. This will be done by throwing a die. The number on the die will indicate the decision round to pick. After that the lottery you chose in that round will be played out. This will be done by throwing two 10-sided dice. The first die gives the first digit, the second die gives the second digit. The lottery will show which numbers correspond to which outcomes (Figure 7).

Example

Look at the Figure 7, it shows the case where round 3 was picked for payoff and the left lottery was chosen in that round. If the dice give a number between 1 and 10, the payoff will be $3. If the dice give a number between 11 and 100, the payoff will be $28.

Appendix B

Chapter 2

B.1 Why Performance Is Not Ability

Consider a simple static optimization model that illustrates why measuring performance does not measure ability, but rather the joint effect of ability and motivation. Suppose that an agent is working on a project, whose success is a Bernoulli random variable with the probability of success p. The agent's preferences over the outcome of the project $X = \{0,1\}$ and exerted effort $E \geqslant 0$ are represented by an additively separable utility function $u(X,E) \in C^2$, strictly increasing in outcome, and strictly decreasing and concave in effort:

$$u(X,E) = \mu X - c(E).$$

The parameter $\mu \geqslant 0$ represents motivation, which captures how valuable the outcome of the project is to the agent. It represents both extrinsic and intrinsic motivation.[1] For the present discussion we do not specify how μ might vary with the monetary reward.[2] The cost of effort function $c(E) \in C^2$ is strictly increasing and convex. It depends on the exerted effort, as opposed to the effective effort. Effective effort includes agent's ability and represents a measure of output.[3] Denoting agent's ability $\alpha \geqslant 0$, the effective effort is $\tilde{E} = f(\alpha,E)$. It increases in both ability and exerted effort, i.e., for any given level of exerted effort, a more able agent produces more effective effort. The probability of success is a function of effective effort $p(\tilde{E}) \in C^2$, strictly increasing and concave. It represents a measure of output or performance, in the sense that if the agent were to repeat N trials of the same project, she would achieve success in pN cases, on average.

The agent's expected utility from the project is

$$U(E) = \mathbb{E}_{\mathrm{E}}\, u(X,E) = \mu\, p(f(\alpha,E)) - c(E), \tag{B.1}$$

[1] We assume that the outcome of the project yields a net benefit to the agent, e.g., if the outcome is inherently unpleasant (negative intrinsic motivation) it is still worth working on it due to a larger positive effect of extrinsic motivation. If the net benefit is non-positive, $\mu \leqslant 0$, the agent would avoid the task completely by exerting zero effort.

[2] While the traditional principal-agent literature would suggest that increasing conditional monetary rewards always increases efforts and performance due to higher motivation, behavioral economics provides evidence that in some important cases extrinsic motivation may crowd out intrinsic motivation and have a negative impact on efforts (Gneezy, Meier, and Rey-Biel, 2011).

[3] In principle, we could avoid introducing effective effort by directly considering how exerted effort affects performance, as for example is done in Lazear (2000), but this additional step makes exposition easier.

where $\mathbb{E}_{\mathrm{E}}$ is the expectation operator, with the subscript reminding that the expectation depends on the chosen level of effort. The expected utility is concave by the assumptions we made so far. Therefore, at the optimal exerted effort E^* the first-order sufficient condition must hold:

$$\mu p'(f(\alpha, E^*)) f_E(\alpha, E^*) - c'(E^*) = 0. \tag{B.2}$$

The probability of success at the optimum will, in general, depend on both parameters, $p^*(\alpha, \mu) = p(f(\alpha, E^*(\alpha, \mu)))$. It will always increase with motivation, but is not guaranteed to increase with ability. To see this, consider that

$$\frac{dp^*}{d\mu} = p' f_E \frac{dE^*}{d\mu},$$

and

$$\operatorname{sgn}\left(\frac{dE^*}{d\mu}\right) = \operatorname{sgn}\left(p' f_E\right) = 1,$$

which follows from the implicit function theorem applied to (B.2), so that $\frac{dp^*}{d\mu} > 0$. On the other hand,

$$\frac{dp^*}{d\alpha} = p' \left(f_\alpha + f_E \frac{dE^*}{d\alpha}\right).$$

Since

$$\operatorname{sgn}\left(\frac{dE^*}{d\alpha}\right) = \operatorname{sgn}\left(\mu p' f_\alpha f_E \left(\frac{f_{\alpha E}}{f_\alpha f_E} + \frac{p''}{p'}\right)\right),$$

the sign of $\frac{dp^*}{d\alpha}$ is undetermined. For example, if the effective effort function is submodular (which implies that $f_{\alpha E} < 0$), the optimal exerted effort will decrease in ability, $\frac{dE^*}{d\alpha} < 0$. If this effect is large enough, it could lead to the negative effect of ability on performance, $\frac{dp^*}{d\alpha} < 0$. Note that performance will not depend on ability at all, if the condition

$$f_\alpha + f_E \frac{dE^*}{d\alpha} = 0$$

holds. For example, consider the following specification:

$$f(\alpha, E) = (\alpha^2 + E^2)^{1/2},\ p(\tilde{E}) = \tilde{E},\ c(E) = E^2,\ \alpha \in [0,1],\ \mu \leqslant 2,\ E \in [0,1],$$

where the restrictions on the parameters and exerted effort ensure that the probability does not exceed 1. Plugging these values into (B.2) yields

$$E^* = \sqrt{(\mu/2)^2 - \alpha^2},\ p^* = \mu/2.$$

Judging the ability of an agent with these preferences based on her performance would be impossible, because the probability of success does not depend on ability at all. The potential positive effect of higher ability on performance is perfectly offset by lower effort, and the performance is determined only by motivation. This case is arguably an extreme one but it illustrates the danger of using performance even as a proxy for ability.

B.2 Analysis of Risk Aversion

In this section, I analyze in more detail how non-parametric risk aversion is measured in the risk task and how this measure of risk aversion is related to a parametric estimate of constant relative risk aversion (CRRA). Consider Figure B.1, which presents a typical round in a risk task. The budget line is defined by two points: M_x and M_y that represent the maximum number of tokens that can be allocated to accounts x and y, respectively. In this case, account y is a cheaper account, since $M_y > M_x$. Point C on the graph corresponds to a point of an equal allocation between accounts. This allocation yields an amount $\bar{x} = \bar{y}$ regardless of an outcome. Only subjects with extremely high degrees of risk aversion would choose this point. Point A represents a risk-neutral allocation, since this allocation yields the highest expected value. Moving from point C to point A represents an increase in the expected value of a lottery while increasing its variance. A moderately risk-averse subject, therefore, would choose an allocation along the line segment AC, which in this example is point B at which x_0 tokens are allocated to account x, and y_0 tokens are allocated to account y. Allocations on the line segment CD are first-order stochastically dominated (FOSD) by all other allocations: moving from C to D results in a decrease in the expected value of a lottery while increasing its variance.

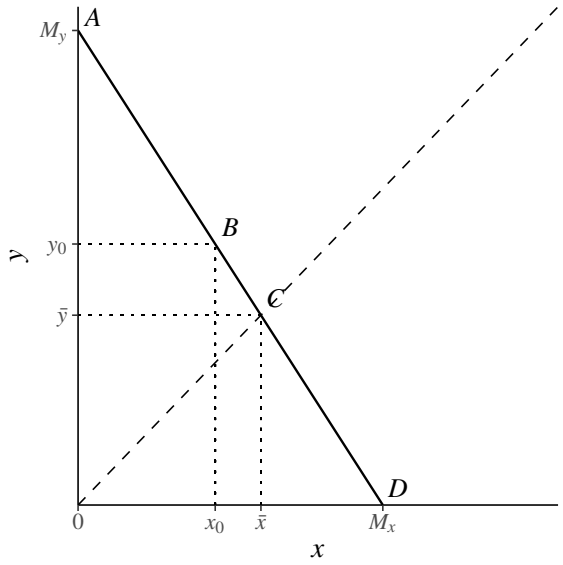


Figure B.1: Geometry of the Risk Task

Intuitively, the closer the chosen allocation point B to the equal-allocation point C, the more risk averse a subject is. On the flip side, the closer point B to a risk-neutral allocation point A, the less risk averse a subject is. Therefore, a subject's risk aversion can be defined as the ratio of the lengths of line segments AB to AC. The resulting ratio should lie in the unit interval for subjects who observe FOSD, with a value of 0 meaning risk neutrality and a value of 1 meaning extreme risk aversion. A value of the ratio greater than 1 implies a violation of FOSD. Practically, the ratio can be found as $x_0/\bar{x}$,[4] where the equal allocation point is given by $\bar{x} = \bar{y} = M_x M_y/(M_x + M_y)$. The

[4]Provided that x is a more expensive account. If x is a cheaper account, the ratio will be defined as $y_0/\bar{y}$.

non-parametric measure of risk aversion for a given subject is then defined as the average of the ratios in each round.

Figure B.2 (left panel) shows the distribution of this measure of risk aversion in the sample. The distribution has a spike around 1, which implies that many subjects have chosen allocations near the equal-allocation point. In fact, 21% of the subjects have estimated risk aversion within a range of 1 ± 0.01. Some subjects can be classified as risk-neutral, which is evident by a small spike around 0. The majority of the subjects, 81%, satisfy FOSD, on average. If one allows for mistakes and classifies subjects as satisfying FOSD with risk aversion below 1.05 instead, then 97% of the subjects will satisfy this criterion.

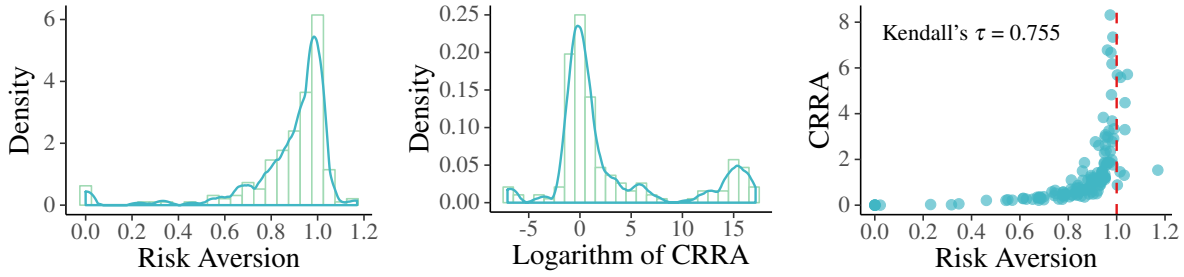


Figure B.2: Risk Aversion in the Sample

To relate the non-parametric measure of risk aversion to a commonly used parametric estimate of CRRA, one would need to estimate the CRRA parameter from the data. Assume that subjects maximize expected utility and have a utility-of-money function of the form $u(x) = x^{1-r}/(1-r)$, where r is the CRRA parameter. One can show that an optimal allocation to account x,[5] as a function of CRRA and design parameters M_x and M_y is given by $x^*(r, M_x, M_y) = aM_xM_y/(M_x + aM_y)$, where $a \equiv (M_x/M_y)^{1/r}$. Higher values of r imply that a is closer to 1, and therefore x^* is closer to $\bar{x}$. On the other hand, lower values of r lead to a chosen allocation being closer to the risk-neutral allocation. Knowing the closed-form solution for the optimal allocation then allows one to estimate r for each subject using data on chosen allocations $\{x\}_i$ and on given budget lines $\{M_x, M_y\}_i$, in each round i.

Figure B.2 (center panel) shows the distribution of the *logarithm* of the estimated CRRA coefficients in the sample. The logarithmic transformation is used so that the picture can fit in the subject with very high CRRA estimates. These are precisely the subjects who chose allocation near the equal-allocation point. The distribution has a spike at 0, which implies a special case of logarithmic utility for many subjects. Overall, the estimated CRRA coefficients are large.

Figure B.2 (right panel) shows the relation between the non-parametric measure of risk aversion and estimated CRRA coefficients.[6] There is a high positive association between the two measures, with Kendall's $\tau = 0.755$ ($p-$value < 0.001). The relationship between the two measures is highly non-linear and is characterized by a convex shape. Around the point where the non-parametric measure of risk aversion approaches 1 (marked by the vertical dashed line) the estimate of CRRA starts to approach infinity, as only subjects with $r = \infty$ would select the equal-allocation point.

[5] One only needs to determine an optimal allocation to one of the accounts, as it exactly pins down an optimal allocation to the remaining account via the budget line equation $y = -(M_y/M_x)x + M_y$.

[6] The graph omits the subjects with CRRA greater than 10 for readability. There are 30% of the subjects with CRRA greater than 10.

B.3 Analysis of CCEI

In this section, I look more closely at the estimates of CCEI in the sample and the associated test power. The CCEI measure introduced by Afriat (1972) quantifies the degree of GARP violations in the sample. CCEI measures by how much the budget lines need to be adjusted to remove all GARP violations. CCEI varies between 0 and 1, with higher values implying that less adjustments are needed and that choices are more consistent with GARP. Lower values of CCEI imply that bigger adjustments are needed and that choices are less consistent with GARP. Figure B.3 (left panel) shows the distribution of the CCEI in the sample, with dashed line showing the median. Subjects show a high degree of consistency with GARP. The median CCEI is 0.982.

An important question in the design of a GARP test is a test power, or one minus the false positive rate. A false positive in the case of GARP would be a situation when a dataset is generated by a subject who's preferences are not consistent with GARP, but who's choices look like they are consistent. To address this issue, I compute the test power using a method by Bronars (1987). I simulate choices for the actual budget lines presented to each subject using 1000 pseudo-subjects who make uniform random choices. I then compute the test power as the proportion of pseudo-subjects who pass the GARP test. I compute the power for three values of the Afriat's adjustment parameter: 0.9, 0.95, and 1. Higher values of the test power imply a stricter test of GARP consistency. Figure B.3 (right panel) shows the distribution of the test power for the three values of Afriat's adjustment parameter. As expected, higher values of the adjustment parameter imply a stricter test. In fact, in case when no adjustments are allowed (the value of the adjustment parameter is 1), almost no pseudo-subjects pass the GARP test. The median test power when the adjustment parameter equals 0.95 is 0.974 and the median test power when the adjustment parameter equals 0.9 is 0.808, which implies that the budget lines used in the experiment in general provide a strict test of GARP consistency.

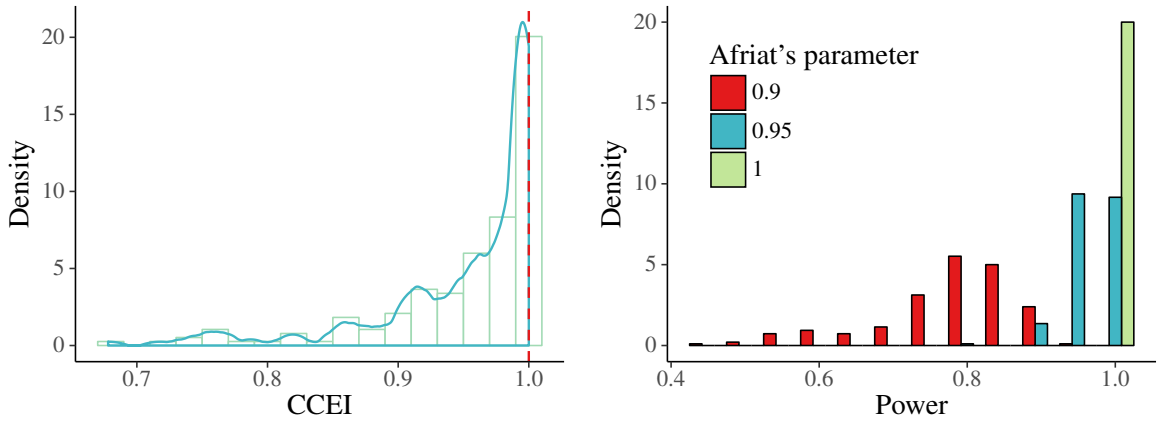


Figure B.3: Distribution of CCEI and Test Power

B.4 Additional Tables

Table B.1: Fractional Regression Results for a Subset of the Data

	(1)	(2)	(3)	(4)	(5)	(6)
Constant	1.980***	1.940***	0.609	1.830***	2.040***	2.310*
	(0.245)	(0.282)	(0.851)	(0.385)	(0.443)	(1.270)
Demographics						
Female	0.273***	0.283***		0.083	0.105	
	(0.087)	(0.079)		(0.140)	(0.125)	
White	0.285*	0.186		−0.017	−0.127	
	(0.152)	(0.125)		(0.234)	(0.188)	
Asian	0.175	0.071		−0.218	−0.372**	
	(0.119)	(0.106)		(0.177)	(0.158)	
Cognitive Abilities						
CRT		0.045	0.042		0.112*	0.057
		(0.037)	(0.040)		(0.058)	(0.062)
GPA		0.053	0.105		0.067	0.052
		(0.088)	(0.096)		(0.138)	(0.144)
Preferences						
Risk Aversion			0.063			−0.103
			(0.192)			(0.285)
CCEI			1.360**			0.310
			(0.627)			(0.990)
Character Skills						
Intellect			0.002			
			(0.008)			
Activity Level			−0.005			−0.008
			(0.009)			(0.015)
Self-Efficacy			0.002			
			(0.012)			
Achievement Striving						−0.004
						(0.014)
Additional Controls	Yes	No	No	Yes	No	No
Observations	99	99	99	99	99	99

Notes: Reports the coefficients of a fractional regression. The data are restricted to a subset of subjects for whom the estimate of non-parametric risk aversion is less than 0.95. Models 1-3 use ability-PoS as a dependent variable, models 4-6 use motivation-PoS as a dependent variable. Standard errors in parenthesis. Omitted categories are female and Black (African American). Additional controls include major, year in school, age, religion, family income, whether a student receives an award, whether a student works, and weekly spending.

*** Significant at the 1 percent level.

** Signficant at the 5 percent level.

* Significant at the 10 percent level.

B.5 Additional Graphs

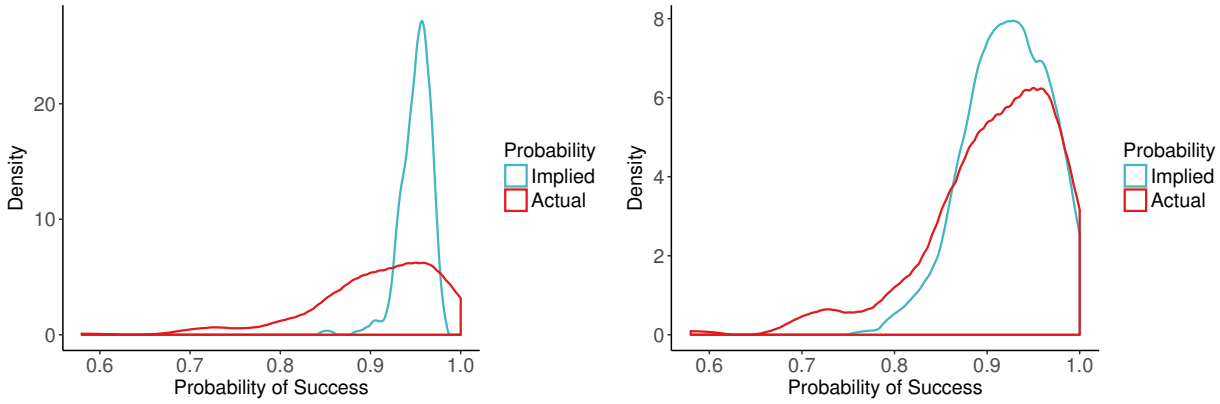


Figure B.4: Distribution of Actual and Implied Probability of Success (Using Means Instead of Medians)

Appendix C

Chapter 3

C.1 Robustness Checks

Here we present additional results derived from alternative assumptions about risk preferences and stochastic choice.

First, we consider an alternative to the EUT, the Rank-Dependent Utility (RDU) model due to Quiggin (1982), which allows for probability weighting. The RDU model has been used extensively in applied and theoretical work. Under this alternative assumption the aggregate utilities of the lotteries are computed as

$$\begin{aligned} U(a;\gamma_u,\gamma_q) = & \\ & = \sum_{i=1}^{k} \left(\omega\left(q_{(1)}(a)+\ldots+q_{(i)}(a);\gamma_q\right) - \omega\left(q_{(1)}(a)+\ldots+q_{(i-1)}(a);\gamma_q\right)\right) \times \\ & \times u\left(x_{(i)}(a);\gamma_u\right), \end{aligned}$$

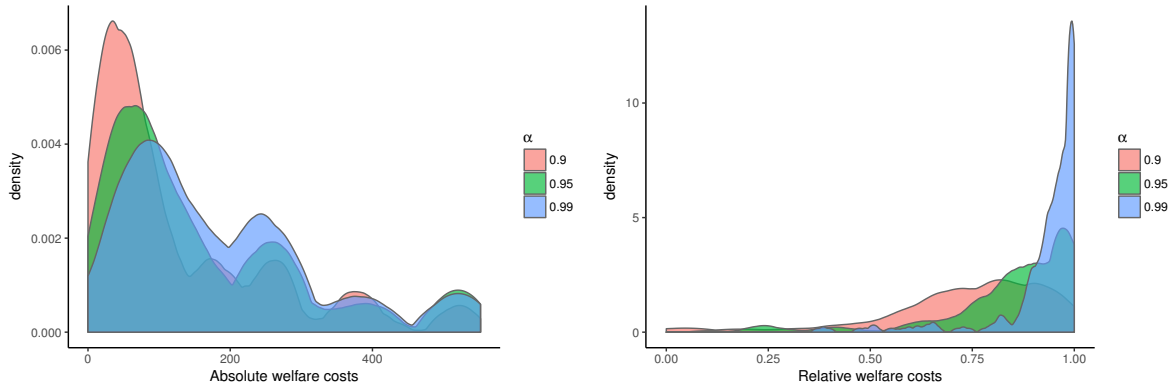
where $\omega : [0,1] \mapsto [0,1]$ is the probability-weighting function, and outcomes are ranked from highest $x_{(1)}$ to lowest $x_{(k)}$, with corresponding probabilities. We assume that ω is the two-parameter (Prelec, 1998) probability weighting function,[1]

$$\omega(q;\gamma_q^1,\gamma_q^2) = \exp(-\gamma_q^2(-\ln q)^{\gamma_q^1}).$$

Figure C.1 shows the calculated absolute and relative welfare costs under the assumption of the RDU model for each individual. Figure C.1 shows that the distributions look very similar to those under EUT, Figure 3.4.

Taking a closer look at the differences between the EUT and RDU-based calculations, we can see from Figure C.2a that the AWC calculated using the EUT model are lower. For $\alpha = 0.9$ the difference in the medians between the AWC calculated using EUT vs. RDU is -59.04 (Wilcoxon signed rank test, p-value < 0.001). The mean of the differences is -84.89 DKK (approximately -13 USD): RDU-based AWC are almost 3 times higher on average.

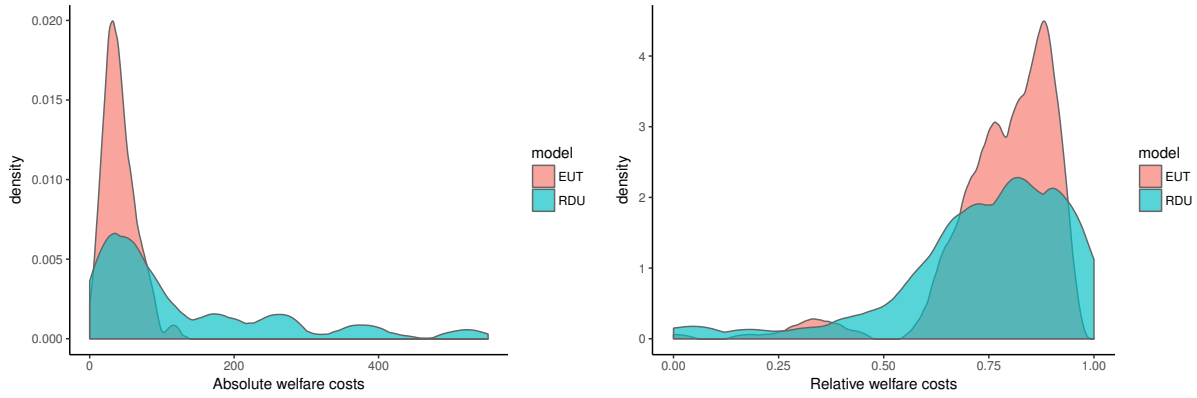
[1] We do not restrict the shape parameter γ_q^1 to in the unit interval, and thus do not impose an inverse-S shape on the probability weighting function.



(a) (b)

Figure C.1: Absolute and Relative Welfare Costs for Three Levels of α, RDU.

The RWC, however, are slightly higher under EUT, as shown in Figure C.2b. The difference in the medians between the RWC calculated using EUT vs. RDU is 0.02 (Wilcoxon signed rank test, p-value $= 0.02$). The mean of the differences is 0.03. The difference in the RWC for RDU and EUT disappears at higher values α, while the difference in the AWC persists. All the other qualitative results on marginal welfare costs, relations between the measures, and observable heterogeneity hold under the RDU assumption.



(a) (b)

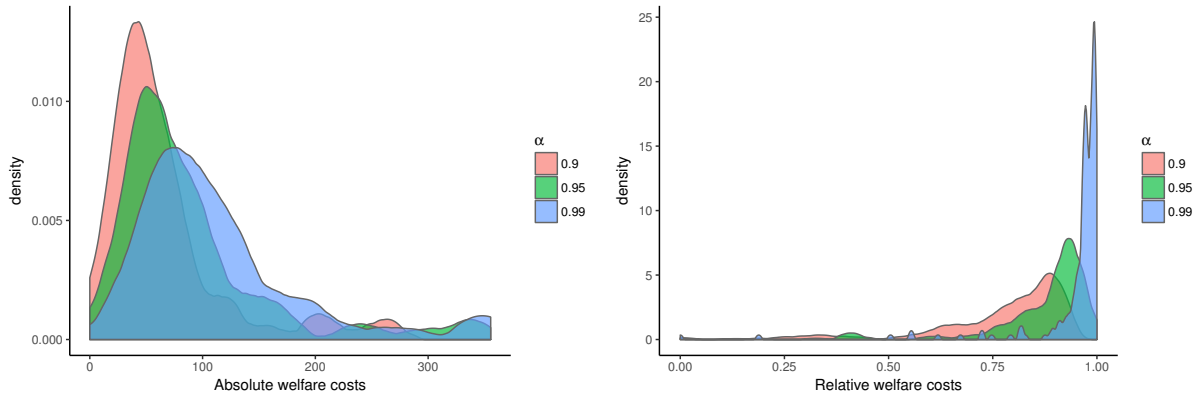
Figure C.2: Absolute and Relative Welfare Costs for EUT vs. RDU, $\alpha = 0.9$.

Second, we consider a different specification for the utility function under EUT, an expo-power (EP) utility, which generalizes the CRRA and CARA utility functions

$$u(x;\gamma_a,\gamma_r) = \frac{1-\exp(-\gamma_a x^{1-\gamma_r})}{\gamma_a},$$

where γ_a and γ_r are the two parameters to be estimated. This specification does not do so well in modeling subjects' risk preferences in our data. For a large (40%) fraction of subjects the estimation procedure yields unreasonably high parameter values, which impedes the calculation of certainty equivalents and welfare costs. We use CRRA specification for these subjects when

presenting the results on Figure C.3. They look very similar to the baseline specification with the CRRA utility function.

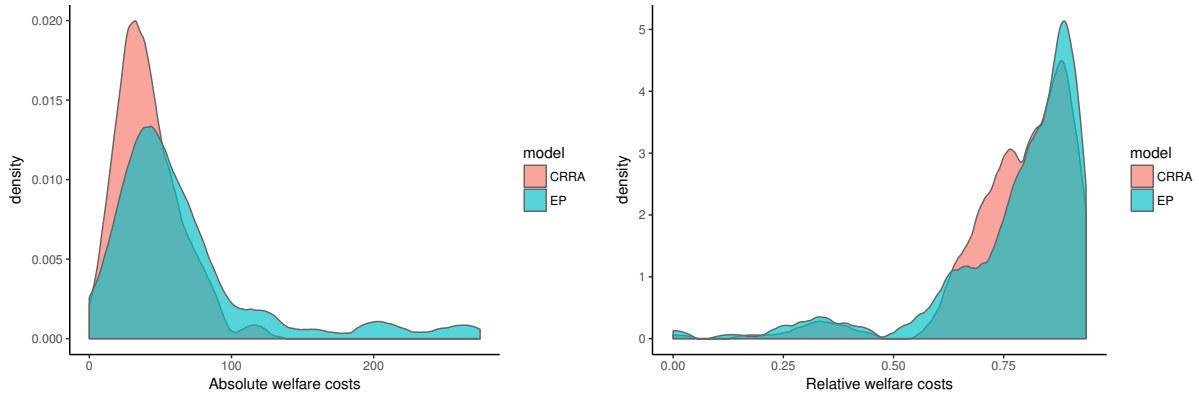


(a) (b)

Figure C.3: Absolute and Relative Welfare Costs for 3 Levels of α, EP.

Looking at the differences between the AWC calculated under the two utility specifications, our baseline specification again provides lower values (see Figure C.4a). The difference in the medians between the AWC calculated using CRRA vs. EP is -16.17 (Wilcoxon signed rank test, p-value < 0.001), for $\alpha = 0.9$. The mean of the differences is -24.7 DKK (approximately -4 USD). The AWC under the EP utility function are roughly 60% higher than in the baseline, which is even higher than in the case of the RDU model as an alternative.

At the same time there are no significant differences in the RWC between the two utility specifications (Wilcoxon signed rank test, p-value = 0.52). The same pattern of results hold for other values of α. Under the EP-utility assumption the marginal welfare costs have a similar shape, but the association between the measures becomes weaker, as do the effects of observable heterogeneity.



(a) (b)

Figure C.4: Absolute and Relative Welfare Costs for CRRA vs. EP, $\alpha = 0.9$.

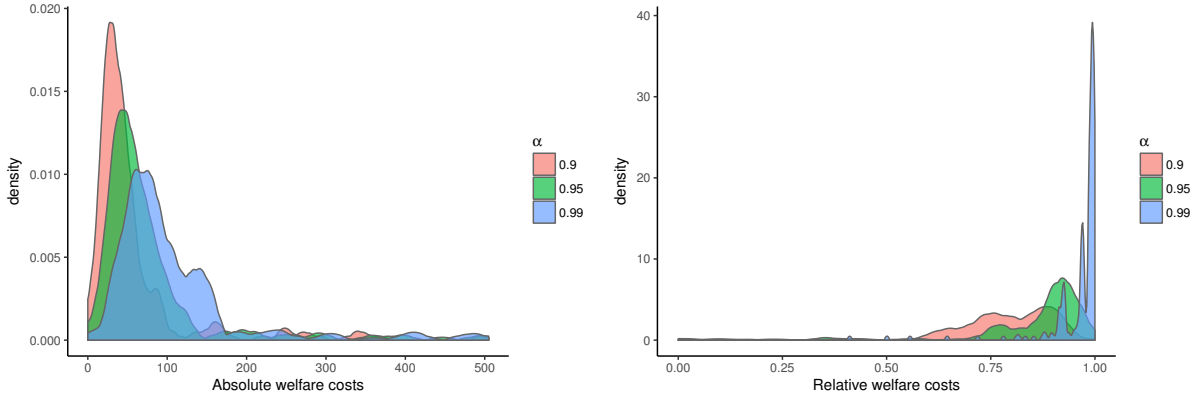
Finally, we look at an alternative stochastic choice specification, the contextual utility model due to Wilcox (2011), which allows for a heterogeneous noise term and preserves the "more risk

averse" relation in the stochastic domain. This specification of noise has been shown by (Wilcox, 2015a) to have good out-of-sample predictive power. Under the assumption of contextual utility the choice probabilities become

$$p(a_2;\gamma,\mu) = \Lambda\left(\frac{U(a_2;\gamma) - U(a_1;\gamma)}{\mu\left(u(x_{(1)};\gamma) - u(x_{(k)};\gamma)\right)}\right),$$

where we drop the index for the decision round, and $p(a_1;\gamma,\mu) = 1 - p(a_2;\gamma,\mu)$. As before, $x_{(1)}$ and $x_{(k)}$ denote the highest and lowest outcomes, but this time they are defined only among the outcomes that occur with positive probabilities, and outcomes are ranked *across both lotteries* in the choice.

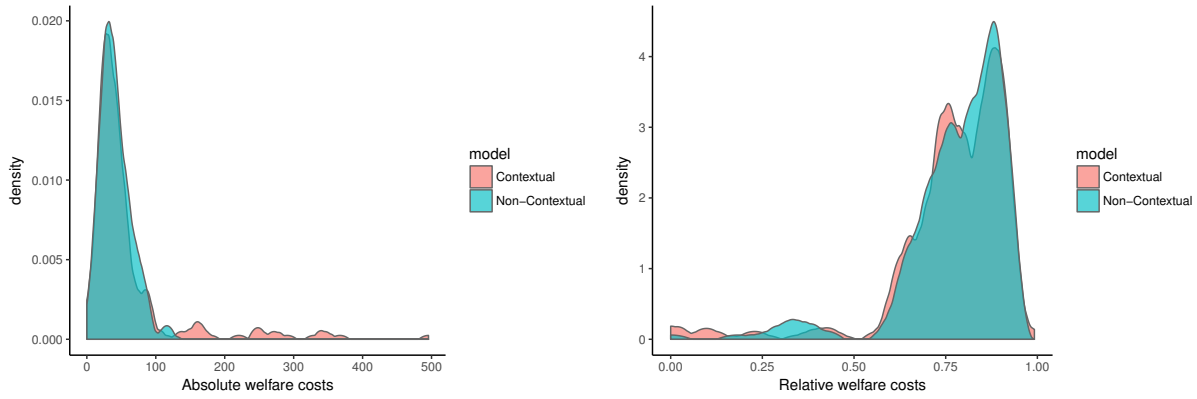
Figure C.5 shows the calculated AWC and RWC under the assumption of contextual utility. These graphs, again, look very similar to those under EUT and no contextual utility (Figure 3.4), except that the right tails in the distributions of the AWC become thicker.



(a) (b)

Figure C.5: Absolute and Relative Welfare Costs for Three Levels of α, Contextual Utility.

Figure C.6 contrasts the AWC and RWC for the baseline and alternative specifications of noise. The densities of the AWC are very much alike, except for a thicker right tail in the case of contextual utility, which leads to higher welfare costs. The difference in the medians between the AWC calculated using non-contextual vs. contextual models is -1.87 (Wilcoxon signed rank test, p-value < 0.001). The mean of the differences is -16.33 DKK (approximately -2 USD). This result is comparable to the non-contextual noise specification with RDU as an alternative. Again, there is no significant difference between the RWC for these two models (Wilcoxon signed rank test, p-value ≈ 0.47). All the results reported for the baseline model hold in the case of contextual utility model as well.



(a) (b)

Figure C.6: Absolute and Relative Welfare Costs for Non-contextual vs. Contextual Utility, $\alpha = 0.9$.

C.2 Proofs

Consider an implicit function $\rho(\mu, \varepsilon) = \alpha$. From the implicit function theorem, it follows that

$$\frac{d\varepsilon}{d\mu} = -\frac{\partial\rho/\partial\mu}{\partial\rho/\partial\varepsilon}.$$

The denominator of this expression is

$$\frac{\partial\rho}{\partial\varepsilon} = \frac{\partial}{\partial\varepsilon}\int_{a_l^*(\varepsilon)}^{a_h^*(\varepsilon)} p(a)da = p(a_h^*(\varepsilon))a_h^{*'}(\varepsilon) - p(a_l^*(\varepsilon))a_l^{*'}(\varepsilon) \geqslant 0,$$

since $a_h^{*'}(\varepsilon) \geqslant 0$, and $a_l^{*'}(\varepsilon) \leqslant 0$.

In order to show the sign of the numerator, we restrict our attention to the binary choice case, since it is the setting of our primary interest. Recall that

$$p(a_2;\gamma,\mu) = \Lambda\left(\frac{U(a_2;\gamma) - U(a_1;\gamma)}{\mu}\right).$$

Then

$$\frac{\partial p(a_2;\gamma,\mu)}{\partial\mu} = \Lambda'\left(\frac{U(a_2;\gamma) - U(a_1;\gamma)}{\mu}\right)(U(a_2;\gamma) - U(a_1;\gamma))(-\mu^2) < 0,$$

since alternative a_2 gives the highest certainty equivalent by our assumption. Therefore,

$$\frac{\partial\rho}{\partial\mu} = \begin{cases} \frac{\partial p(a_2;\gamma,\mu)}{\partial\mu}, & \varepsilon < \Delta m, \\ 0, & \varepsilon \geqslant \Delta m, \end{cases}$$

so that $\partial\rho/\partial\mu \leqslant 0$. Together the two results imply that $d\varepsilon/d\mu \geqslant 0$.

C.3 Additional Tables

Table C.1: The Battery of Lotteries

ID	Lp1	La1	Lp2	La2	Lp3	La3	Lp4	La4	Rp1	Ra1	Rp2	Ra2	Rp3	Ra3	Rp4	Ra4
1	0.50	300	0	900	0.50	1,500	0	0	0.10	300	0.80	900	0.10	1,500	0	0
2	0.50	300	0	900	0.50	1,500	0	0	0	300	1	900	0	1,500	0	0
3	0.10	300	0.80	900	0.10	1,500	0	0	0	300	1	900	0	1,500	0	0
4	0.70	300	0	900	0.30	1,500	0	0	0.50	300	0.40	900	0.10	1,500	0	0
5	0.70	300	0	900	0.30	1,500	0	0	0.40	300	0.60	900	0	1,500	0	0
6	0.50	300	0.40	900	0.10	1,500	0	0	0.40	300	0.60	900	0	1,500	0	0
7	0.40	300	0	900	0.60	1,500	0	0	0.10	300	0.75	900	0.15	1,500	0	0
8	0.40	300	0	900	0.60	1,500	0	0	0	300	1	900	0	1,500	0	0
9	0.30	300	0	900	0.70	1,500	0	0	0.15	300	0.25	900	0.60	1,500	0	0
10	0.10	300	0.75	900	0.15	1,500	0	0	0	300	1	900	0	1,500	0	0
11	0.70	300	0	900	0.30	1,500	0	0	0.60	300	0.25	900	0.15	1,500	0	0
12	0.70	300	0	900	0.30	1,500	0	0	0.50	300	0.50	900	0	1,500	0	0
13	0.60	300	0.25	900	0.15	1,500	0	0	0.50	300	0.50	900	0	1,500	0	0
14	0.40	300	0	900	0.60	1,500	0	0	0.20	300	0.60	900	0.20	1,500	0	0
15	0.40	300	0	900	0.60	1,500	0	0	0.10	300	0.90	900	0	1,500	0	0
16	0.20	300	0.60	900	0.20	1,500	0	0	0.10	300	0.90	900	0	1,500	0	0
17	0.60	300	0	900	0.40	1,500	0	0	0.50	300	0.30	900	0.20	1,500	0	0
18	0.30	300	0	900	0.70	1,500	0	0	0	300	0.50	900	0.50	1,500	0	0
19	0.60	300	0	900	0.40	1,500	0	0	0.40	300	0.60	900	0	1,500	0	0
20	0.50	300	0.30	900	0.20	1,500	0	0	0.40	300	0.60	900	0	1,500	0	0
21	0.25	300	0	900	0.75	1,500	0	0	0.10	300	0.60	900	0.30	1,500	0	0
22	0.25	300	0	900	0.75	1,500	0	0	0	300	1	900	0	1,500	0	0
23	0.10	300	0.60	900	0.30	1,500	0	0	0	300	1	900	0	1,500	0	0
24	0.50	300	0.20	900	0.30	1,500	0	0	0.40	300	0.60	900	0	1,500	0	0
25	0.55	300	0	900	0.45	1,500	0	0	0.40	300	0.60	900	0	1,500	0	0
26	0.55	300	0	900	0.45	1,500	0	0	0.50	300	0.20	900	0.30	1,500	0	0
27	0.15	300	0.25	900	0.60	1,500	0	0	0	300	0.50	900	0.50	1,500	0	0
28	0.15	300	0.75	900	0.10	1,500	0	0	0	300	1	900	0	1,500	0	0
29	0.60	300	0	900	0.40	1,500	0	0	0	300	1	900	0	1,500	0	0
30	0.60	300	0	900	0.40	1,500	0	0	0.15	300	0.75	900	0.10	1,500	0	0
31	0.55	90	0.25	1,080	0.20	1,260	0	0	0.55	90	0.25	810	0.20	1,620	0	0
32	0.40	540	0.40	450	0.20	1,080	0	0	0.40	540	0.40	270	0.20	1,350	0	0
33	0.40	990	0.40	450	0.20	1,080	0	0	0.40	990	0.40	270	0.20	1,350	0	0
34	0.40	1,440	0.40	450	0.20	1,080	0	0	0.40	1,440	0.40	270	0.20	1,350	0	0
35	0.70	450	0.10	990	0.20	1,890	0	0	0.70	450	0.10	630	0.20	2,250	0	0
36	0.70	1,080	0.10	990	0.20	1,890	0	0	0.70	1,080	0.10	630	0.20	2,250	0	0
37	0.70	1,710	0.10	990	0.20	1,890	0	0	0.70	1,710	0.10	630	0.20	2,250	0	0
38	0.70	2,340	0.10	990	0.20	1,890	0	0	0.70	2,340	0.10	630	0.20	2,250	0	0
39	0.50	0	0.10	360	0.40	360	0	0	0.50	0	0.10	0	0.40	540	0	0
40	0.50	360	0.10	360	0.40	360	0	0	0.50	360	0.10	0	0.40	540	0	0
41	0.50	720	0.10	360	0.40	360	0	0	0.50	720	0.10	0	0.40	540	0	0
42	0.55	630	0.25	1,080	0.20	1,260	0	0	0.55	630	0.25	810	0.20	1,620	0	0
43	0.50	1,080	0.10	360	0.40	360	0	0	0.50	1,080	0.10	0	0.40	540	0	0
44	0.50	360	0.10	720	0.40	720	0	0	0.50	360	0.10	360	0.40	900	0	0
45	0.50	720	0.10	720	0.40	720	0	0	0.50	720	0.10	360	0.40	900	0	0
46	0.50	1,080	0.10	720	0.40	720	0	0	0.50	1,080	0.10	360	0.40	900	0	0
47	0.50	1,440	0.10	720	0.40	720	0	0	0.50	1,440	0.10	360	0.40	900	0	0
48	0.55	1,170	0.25	1,080	0.20	1,260	0	0	0.55	1,170	0.25	810	0.20	1,620	0	0
49	0.55	1,710	0.25	1,080	0.20	1,260	0	0	0.55	1,710	0.25	810	0.20	1,620	0	0
50	0.65	90	0.20	630	0.15	990	0	0	0.65	90	0.20	540	0.15	1,080	0	0
51	0.65	450	0.20	630	0.15	990	0	0	0.65	450	0.20	540	0.15	1,080	0	0
52	0.65	810	0.20	630	0.15	990	0	0	0.65	810	0.20	540	0.15	1,080	0	0
53	0.65	1,170	0.20	630	0.15	990	0	0	0.65	1,170	0.20	540	0.15	1,080	0	0
54	0.40	90	0.40	450	0.20	1,080	0	0	0.40	90	0.40	270	0.20	1,350	0	0
55	0	0	0	0	0	0	1	800	0	0	0	0	0.50	650	0.50	960
56	0	0	0	0	0	0	1	850	0	0	0	0	0.50	770	0.50	940
57	0	0	0	0	0	0	1	300	0	0	0	0	0.50	150	0.50	460
58	0	0	0	0	0	0	1	1,300	0	0	0	0	0.50	1,150	0.50	1,460
59	0	0	0	0	0	0	1	1,350	0	0	0	0	0.50	1,270	0.50	1,440
60	0	0	0	0	0	0	1	150	0	0	0	0	0.50	70	0.50	240

Notes. The columns are coded as follows: “L" and “R" denote left and right lottery, “a" denotes amounts (in DKK) and “p" denotes probabilities. The amounts in the table are baseline amounts. In addition to these amounts, 1.5x and 2x amounts were used. The subjects were randomized across the baseline, 1.5x and 2x amounts.

Bibliography

Abeler J, Falk A, Goette L, Huffman D (2011). "Reference Points and Effort Provision." *American Economic Review*, **101**(2), 470–92.

Afriat SN (1972). "Efficiency Estimation of Production Function." *International Economic Review*, **13**(3), 568–98.

Agarwal S, Mazumder B (2013). "Cognitive Abilities and Household Financial Decision Making." *American Economic Journal: Applied Economics*, **5**(1), 193–207.

Agranov M, Ortoleva P (2017). "Stochastic Choice and Preferences for Randomization." *Journal of Political Economy*, **125**(1), 40–68.

Alekseev A, Charness G, Gneezy U (2017). "Experimental Methods: When and Why Contextual Instructions Are Important." *Journal of Economic Behavior & Organization*, **134**, 48–59.

Alter AL, Oppenheimer DM, Epley N, Eyre RN (2007). "Overcoming Intuition: Metacognitive Difficulty Activates Analytic Reasoning." *Journal of Experimental Psychology: General*, **136**(4), 569–576.

Amaresha AC, Danivas V, Shivakumar V, Agarwal SM, Kalmady SV, Narayanaswamy JC, Venkatasubramanian G (2014). "Clinical Correlates of Parametric Digit-Symbol Substitution Test in Schizophrenia." *Asian Journal of Psychiatry*, **10**, 45–50.

Andersen S, Harrison GW, Lau MI, Rutström EE (2008). "Eliciting Risk and Time Preferences." *Econometrica*, **76**(3), 583–618.

Andersen S, Harrison GW, Lau MI, Rutström EE (2018). "Multiattribute Utility Theory, Intertemporal Utility, And Correlation Aversion." *International Economic Review*, **59**(2), 537–555.

Apesteguia J, Ballester MA, Lu J (2017). "Single-Crossing Random Utility Models." *Econometrica*, **85**(2), 661–674. ISSN 1468-0262.

Ariely D, Gneezy U, Loewenstein G, Mazar N (2009). "Large Stakes and Big Mistakes." *Review of Economic Studies*, **76**(2), 451–469.

Ballinger TP, Wilcox NT (1997). "Decisions, Error and Heterogeneity." *Economic Journal*, **107**(443), 1090–1105.

Balzer WK, Doherty ME, O'Connor R (1989). "Effects of Cognitive Feedback on Performance." *Psychological Bulletin*, **106**(3), 410.

Barrick MR, Mount MK (1991). "The Big Five Personality Dimensions and Job Performance: A Meta-Analysis." *Personnel Psychology*, **44**(1), 1–26.

Beatty TKM, Crawford IA (2011). "How Demanding Is the Revealed Preference Approach to Demand?" *American Economic Review*, **101**(6), 2782–95.

Bland JR (2018). "The Cost of Being Behavioral in Risky Choice Experiments." *Working paper*, The University of Toledo, Department of Economics.

Blanes i Vidal J, Nossol M (2011). "Tournaments Without Prizes: Evidence From Personnel Records." *Management Science*, **57**(10), 1721–1736.

Blavatskyy PR (2014). "Stronger Utility." *Theory and Decision*, **76**(2), 265–286.

Blow L, Browning M, Crawford I (2008). "Revealed Preference Analysis of Characteristics Models." *Review of Economic Studies*, **75**(2), 371–389.

Borghans L, Duckworth AL, Heckman JJ, Ter Weel B (2008). "The Economics and Psychology of Personality Traits." *Journal of Human Resources*, **43**(4), 972–1059.

Borghans L, Golsteyn BH, Heckman J, Humphries JE (2011). "Identification Problems in Personality Psychology." *Personality and Individual Differences*, **51**(3), 315–320.

Borghans L, Meijers H, Ter Weel B (2013). "The Importance of Intrinsic and Extrinsic Motivation for Measuring IQ." *Economics of Education Review*, **34**, 17–28.

Brehm JW, Self EA (1989). "The Intensity of Motivation." *Annual Review of Psychology*, **40**, 109–131.

Bremzen A, Khokhlova E, Suvorov A, Van de Ven J (2015). "Bad News: An Experimental Study on the Informational Effects of Rewards." *Review of Economics and Statistics*, **97**(1), 55–70.

Brockner J, Grover S, Reed TF, Dewitt RL (1992). "Layoffs, Job Insecurity, and Survivors' Work Effort: Evidence of an Inverted-U Relationship." *Academy of Management Journal*, **35**(2), 413–425.

Bronars SG (1987). "The Power of Nonparametric Tests of Preference Maximization." *Econometrica*, **55**(3), 693–698.

Brüggen A, Strobel M (2007). "Real Effort Versus Chosen Effort in Experiments." *Economics Letters*, **96**(2), 232–236.

Bruhin A, Fehr-Duda H, Epper T (2010). "Risk and Rationality: Uncovering Heterogeneity in Probability Distortion." *Econometrica*, **78**(4), 1375–1412.

Bull C, Schotter A, Weigelt K (1987). "Tournaments and Piece Rates: An Experimental Study." *Journal of Political Economy*, **95**(1), 1–33.

Busemeyer JR, Townsend JT (1993). "Decision Field Theory: A Dynamic-Cognitive Approach to Decision Making in an Uncertain Environment." *Psychological review*, **100**(3), 432.

Camerer CF (1989). "An Experimental Test of Several Generalized Utility Theories." *Journal of Risk and Uncertainty*, **2**(1), 61–104.

Campbell DJ, Ilgen DR (1976). "Additive Effects of Task Difficulty and Goal Setting on Subsequent Task Performance." *Journal of Applied Psychology*, **61**(3), 319.

Caplin A, Dean M (2015). "Revealed Preference, Rational Inattention, and Costly Information Acquisition." *American Economic Review*, **105**(7), 2183–2203.

Cattell RB (1971). *Abilities: Their Structure, Growth, and Action*. Boston: Houghton Mifflin.

Charness G, Cobo-Reyes R, Jiménez N, Lacomba JA, Lagos F (2012). "The Hidden Advantage of Delegation: Pareto Improvements in a Gift Exchange Game." *American Economic Review*, **102**(5), 2358–79.

Cheremukhin A, Popova A, Tutino A (2015). "A Theory of Discrete Choice with Information Costs." *Journal of Economic Behavior & Organization*, **113**, 34–50.

Chetty R, Friedman JN, Hilger N, Saez E, Schanzenbach DW, Yagan D (2011). "How Does Your Kindergarten Classroom Affect Your Earnings? Evidence from Project Star." *The Quarterly Journal of Economics*, **126**(4), 1593–1660.

Choi S, Fisman R, Gale D, Kariv S (2007). "Consistency and Heterogeneity of Individual Behavior under Uncertainty." *American Economic Review*, **97**(5), 1921–1938.

Choi S, Kariv S, Müller W, Silverman D (2014). "Who Is (More) Rational?" *American Economic Review*, **104**(6), 1518–1550.

Cohn A, Fehr E, Goette L (2014). "Fair Wages and Effort Provision: Combining Evidence From a Choice Experiment and a Field Experiment." *Management Science*, **61**(8), 1777–1794.

Corgnet B, Gómez-Miñambres J, Hernán-González R (2015). "Goal Setting and Monetary Incentives: When Large Stakes Are Not Enough." *Management Science*, **61**(12), 2926–2944.

Cox JC, Sadiraj V (2008). "Risky Decisions in the Large and in the Small: Theory and Experiment." In JC Cox, GW Harrison (eds.), "Risk Aversion in Experiments," volume 12 of *Research in Experimental Economics*, pp. 9–39. Bingley, UK: Emerald.

Cox JC, Sadiraj V, Schmidt U (2015). "Paradoxes and Mechanisms for Choice Under Risk." *Experimental Economics*, **18**(2), 215–250.

Deary IJ, Der G (2005). "Reaction Time, Age, and Cognitive Ability: Longitudinal Findings from Age 16 to 63 Years in Representative Population Samples." *Aging, Neuropsychology, and cognition*, **12**(2), 187–215.

Dennett D (1987). *The Intentional Stance*. Cambridge, MA: MIT Press.

Dohmen T, Falk A, Huffman D, Sunde U (2010). "Are Risk Aversion and Impatience Related to Cognitive Ability?" *American Economic Review*, **100**(3), 1238–60.

Duckworth AL, Quinn PD, Lynam DR, Loeber R, Stouthamer-Loeber M (2011). "Role of Test Motivation in Intelligence Testing." *Proceedings of the National Academy of Sciences*, **108**(19), 7716–7720.

Ewers M, Zimmermann F (2015). "Image and Misreporting." *Journal of the European Economic Association*, **13**(2), 363–380.

Falk A, Becker A, Dohmen TJ, Huffman D, Sunde U (2016). "The Preference Survey Module: A Validated Instrument for Measuring Risk, Time, and Social Preferences." *Discussion Paper 9674*, Institute for the Study of Labor (IZA).

Fechner GT (1860). *Elements of Psychophysics*. Amsterdam: Bonset.

Fehr E, Gächter S, Kirchsteiger G (1997). "Reciprocity as a Contract Enforcement Device: Experimental Evidence." *Econometrica*, **65**(4), 833–860.

Filiz-Ozbay E, Guryan J, Hyndman K, Kearney MS, Ozbay EY (2015). "Do Lottery Payments Induce Savings Behavior: Evidence from the Lab." *Journal of Public Economics*, **126**, 1–24.

Fischbacher U (2007). "z-Tree: Zurich Toolbox for Ready-Made Economic Experiments." *Experimental Economics*, **10**(2), 171–178.

Frederick S (2005). "Cognitive Reflection and Decision Making." *Journal of Economic Perspectives*, **19**(4), 25–42.

Fryer Jr RG, Levitt SD, List J, Sadoff S (2012). "Enhancing the Efficacy of Teacher Incentives Through Loss Aversion: A Field Experiment." *Working Paper 18237*, National Bureau of Economic Research.

Genakos C, Pagliero M, Garbi E (2015). "When Pressure Sinks Performance: Evidence From Diving Competitions." *Economics Letters*, **132**, 5–8.

Gendolla GH, Richter M, Silvia PJ (2008). "Self-Focus and Task Difficulty Effects on Effort-Related Cardiovascular Reactivity." *Psychophysiology*, **45**(4), 653–662.

Gendolla GH, Wright RA, Richter M (2012). "Effort Intensity: Some Insights From the Cardiovascular System." In R Ryan (ed.), *The Oxford Handbook of Human Motivation*, pp. 420–438. Oxford University Press.

Gill D, Prowse V (2016). "Cognitive Ability, Character Skills, and Learning to Play Equilibrium: A Level-k Analysis." *Journal of Political Economy*, **124**(6), 1619–1676.

Gneezy U, Kajackaite A, Sobel J (2018). "Lying Aversion and the Size of the Lie." *American Economic Review*, **108**(2), 419–53.

Gneezy U, List JA (2006). "Putting Behavioral Economics to Work: Testing for Gift Exchange in Labor Markets Using Field Experiments." *Econometrica*, **74**(5), 1365–1384.

Gneezy U, Meier S, Rey-Biel P (2011). "When and Why Incentives (Don't) Work to Modify Behavior." *Journal of Economic Perspectives*, **25**(4), 191–209.

Gneezy U, Rustichini A (2000). "Pay Enough Or Don't Pay At All." *Quarterly Journal of Economics*, **115**(3), 791–810.

Goerg SJ, Kube S (2012). "Goals (th)at Work." *Working Paper Bonn 2012/19*, Max Planck Institute for Research on Collective Goods.

Gold JI, Shadlen MN (2007). "The Neural Basis Of Decision Making." *Annual Review of Neuroscience*, **30**, 535–574.

Gul F, Natenzon P, Pesendorfer W (2014). "Random Choice as Behavioral Optimization." *Econometrica*, **82**(5), 1873–1912.

Gul F, Pesendorfer W (2006). "Random Expected Utility." *Econometrica*, **74**(1), 121–146.

Harless DW, Camerer CF (1994). "The Predictive Utility of Generalized Expected Utility Theories." *Econometrica*, **62**(6), 1251–1289.

Harrison GW (1989). "Theory and Misbehavior of First-Price Auctions." *American Economic Review*, **79**(4), 749–762.

Harrison GW (1992). "Theory and Misbehavior of First-Price Auctions: Reply." *American Economic Review*, **82**(5), 1426–1443.

Harrison GW (1994). "Expected Utility Theory and the Experimentalists." *Empirical Economics*, **19**, 223–253.

Harrison GW, Jessen LJ, Lau M, Ross D (2018). "Disordered Gambling Prevalence: Methodological Innovations in a General Danish Population Survey." *Journal of Gambling Studies*, **34**(1), 225–253.

Harrison GW, List JA (2004). "Field Experiments." *Journal of Economic Literature*, **42**(4), 1009–1055.

Harrison GW, Morgan P (1990). "Search Intensity in Experiments." *Economic Journal*, **100**(401), 478–486.

Harrison GW, Ng JM (2016). "Evaluating the Expected Welfare Gain from Insurance." *Journal of Risk and Insurance*, **83**(1), 91–120.

Harrison GW, Ross D (2018). "Varieties of Paternalism and the Heterogeneity of Utility Structures." *Journal of Economic Methodology*, **25**(1), 42–67.

Harrison GW, Rutström EE (2008). "Risk Aversion in the Laboratory." In JC Cox and GW Harrison (eds.), *Risk Aversion in Experiments*. Bingley, UK: Emerald, Research in Experimental Economics, Volume 12.

Harrison GW, Swarthout JT (2014). "Experimental Payment Protocols and the Bipolar Behaviorist." *Theory and Decision*, **77**(3), 423–438.

Heath C, Larrick RP, Wu G (1999). "Goals as Reference Points." *Cognitive Psychology*, **38**(1), 79–109.

Heckman J, Pinto R, Savelyev P (2013). "Understanding the Mechanisms Through Which an Influential Early Childhood Program Boosted Adult Outcomes." *American Economic Review*, **103**(6), 2052–2086.

Heckman JJ, Stixrud J, Urzua S (2006). "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior." *Journal of Labor Economics*, **24**(3), 411–482.

Hennig-Schmidt H, Sadrieh A, Rockenbach B (2010). "In Search of Workers' Real Effort Reciprocity—a Field and a Laboratory Experiment." *Journal of the European Economic Association*, **8**(4), 817–837.

Hertwig R, Barron G, Weber EU, Erev I (2004). "Decisions From Experience and the Effect of Rare Events in Risky Choice." *Psychological Science*, **15**(8), 534–539.

Hertwig R, Erev I (2009). "The Description–Experience Gap in Risky Choice." *Trends in Cognitive Sciences*, **13**(12), 517–523.

Hey JD (2001). "Does Repetition Improve Consistency?" *Experimental Economics*, **4**(1), 5–54.

Hey JD, Orme C (1994). "Investigating Generalizations of Expected Utility Theory Using Experimental Data." *Econometrica*, **62**(6), 1291–1326.

Hickman DC, Metz NE (2015). "The Impact of Pressure on Performance: Evidence From the Pga Tour." *Journal of Economic Behavior & Organization*, **116**, 319–330.

Hoch SJ, Loewenstein GF (1989). "Outcome Feedback: Hindsight and Information." *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **15**(4), 605.

Holmström B (2017). "Pay for Performance and Beyond." *American Economic Review*, **107**(7), 1753–77.

Holt CA, Laury SK (2002). "Risk Aversion and Incentive Effects." *American Economic Review*, **92**(5), 1644–1655.

Hossain T, List J (2012). "The Behavioralist Visits the Factory: Increasing Productivity Using Simple Framing Manipulations." *Management Science*, **58**(12), 2151–2167.

Hunt T, Asplund C (2010). "Concussion Assessment and Management." *Clinics in Sports Medicine*, **29**(1), 5–17.

Infante G, Lecouteux G, Sugden R (2016). "Preference Purification and the Inner Rational Agent: A Critique of the Conventional Wisdom of Behavioural Welfare Economics." *Journal of Economic Methodology*, **23**(1), 1–25.

Jayaraman R, Ray D, de Vèricourt F (2016). "Anatomy of a Contract Change." *American Economic Review*, **106**(2), 316–58.

Joy S, Kaplan E, Fein D (2004). "Speed and Memory in the WAIS-III Digit Symbol-Coding Subtest Across the Adult Lifespan." *Archives of Clinical Neuropsychology*, **19**(6), 759–767.

Kieburtz K (1996). "Unified Huntington's Disease Rating Scale: Reliability and Consistency." *Movement Disorders*, **11**(2), 136–142.

Labroo AA, Kim S (2009). "The 'Instrumentality' Heuristic Why Metacognitive Difficulty Is Desirable During Goal Pursuit." *Psychological Science*, **20**(1), 127–134.

Lafont S, Marin-Lamellet C, Paire-Ficout L, Thomas-Anterion C, Laurent B, Fabrigoule C (2010). "The Wechsler Digit Symbol Substitution Test as the Best Indicator of the Risk of Impaired Driving in Alzheimer Disease and Normal Aging." *Dementia and Geriatric Cognitive Disorders*, **29**(2), 154–163.

Lazear EP (2000). "Performance Pay and Productivity." *American Economic Review*, **90**(5), 1346–1361.

Lee MLT, Whitmore G (2006). "Threshold Regression for Survival Analysis: Modeling Event Times by a Stochastic Process Reaching a Boundary." *Statistical Science*, pp. 501–513.

Loomes G, Sugden R (1995). "Incorporating a Stochastic Element into Decision Theories." *European Economic Review*, **39**(3), 641–648.

Loomes G, Sugden R (1998). "Testing Different Stochastic Specifications of Risky Choice." *Economica*, **65**(260), 581–598.

Luce RD (1959). *Individual Choice Behavior: A Theoretical Analysis*. Wiley, New York.

Marschak J (1960). "Binary-Choice Constraints and Random Utility Indicators." In K Arrow (ed.), "Stanford Symposium on Mathematical Methods in the Social Sciences," pp. 312–29. Stanford, CA: Stanford University Press.

Matêjka F, McKay A (2015). "Rational Inattention to Discrete Choices: A New Foundation for the Multinomial Logit Model." *American Economic Review*, **105**(1), 272–98.

McDaniel TM, Rutström EE (2001). "Decision Making Costs and Problem Solving Performance." *Experimental Economics*, **4**(2), 145–161.

McFadden DL (1976). "Quantal Choice Analaysis: A Survey." In SV Berg (ed.), "Annals of Economic and Social Measurement," volume 5, pp. 363–390. NBER.

McFadden DL (2001). "Economic Choices." *American Economic Review*, **91**(3), 351–378.

Monroe BA (2017). *Stochastic Models in Experimental Economics*. Ph.D. thesis, University of Cape Town.

Murnane R, Willett JB, Levy F (1995). "The Growing Importance of Cognitive Skills in Wage Determination." *The Review of Economics and Statistics*, **77**(2), 251–66.

Nilsson JP (2017). "Alcohol Availability, Prenatal Conditions, and Long-Term Economic Outcomes." *Journal of Political Economy*, **125**(4), 1149–1207.

Nogee P, Mosteller F (1951). "An Experimental Measure of Utility." *Journal of Political Economy*, **59**, 371–404.

Noussair CN, Trautmann ST, van de Kuilen G (2014). "Higher Order Risk Attitudes, Demographics, and Financial Decisions." *The Review of Economic Studies*, **81**(1), 325–355.

Ofek E, Yildiz M, Haruvy E (2007). "The Impact of Prior Decisions on Subsequent Valuations in a Costly Contemplation Model." *Management Science*, **53**(8), 1217–1233.

Papke LE, Wooldridge JM (1996). "Econometric Methods for Fractional Response Variables with an Application to 401 (K) Plan Participation Rates." *Journal of Applied Econometrics*, **11**(6), 619–32.

Prelec D (1998). "The Probability Weighting Function." *Econometrica*, **66**(3), 497–528.

Proust-Lima C, Amieva H, Dartigues JF, Jacqmin-Gadda H (2007). "Sensitivity of Four Psychometric Tests to Measure Cognitive Changes in Brain Aging-Population–Based Studies." *American Journal of Epidemiology*, **165**(3), 344–350.

Quiggin J (1982). "A Theory of Anticipated Utility." *Journal of Economic Behavior & Organization*, **3**(4), 323–343.

Ratcliff R (1978). "A Theory of Memory Retrieval." *Psychological Review*, **85**(2), 59.

Richter M (2015). "Goal Pursuit and Energy Conservation: Energy Investment Increases With Task Demand But Does Not Equal It." *Motivation and Emotion*, **39**(1), 25–33.

Richter M, Friedrich A, Gendolla GH (2008). "Task Difficulty Effects on Cardiac Activity." *Psychophysiology*, **45**(5), 869–875.

Ross D (2014). *Philosophy of Economics*. London: Palgrave Macmillan.

Segal C (2012). "Working When No One Is Watching: Motivation, Test Scores, and Economic Success." *Management Science*, **58**(8), 1438–1457.

Smith TW, Baldwin M, Christensen AJ (1990). "Interpersonal Influence as Active Coping: Effects of Task Difficulty on Cardiovascular Reactivity." *Psychophysiology*, **27**(4), 429–437.

Smithers S (2015). "Goals, Motivation and Gender." *Economics Letters*, **131**, 75–77.

Spiliopoulos L, Ortmann A (2018). "The BCD of Response Time Analysis in Experimental Economics." *Experimental Economics*, **21**(2), 383–433.

Starmer C, Sugden R (1989). "Probability and Juxtaposition Effects: An Experimental Investigation of the Common Ratio Effect." *Journal of Risk and Uncertainty*, **2**(2), 159–78.

Stott HP (2006). "Cumulative Prospect Theory's Functional Menagerie." *Journal of Risk and Uncertainty*, **32**(2), 101–130.

Swait J, Marley AAJ (2013). "Probabilistic Choice (Models) as a Result of Balancing Multiple Goals." *Journal of Mathematical Psychology*, **57**(1–2), 1–14.

Thurstone LL (1927). "A Law Of Comparative Judgment." *Psychological review*, **34**(4), 266–270.

Tversky A (1969). "Intransitivity of Preferences." *Psychological review*, **76**(1), 31.

Van de Kuilen G (2009). "Subjective Probability Weighting and the Discovered Preference Hypothesis." *Theory and Decision*, **67**(1), 1–22.

Vandegrift D, Brown P (2003). "Task Difficulty, Incentive Effects, and the Selection of High-Variance Strategies: An Experimental Examination of Tournament Behavior." *Labour Economics*, **10**(4), 481–497.

Vernon PA (1983). "Speed of Information Processing and General Intelligence." *Intelligence*, **7**(1), 53–70.

von Gaudecker HM, van Soest A, Wengstrom E (2011). "Heterogeneity in Risky Choice Behavior in a Broad Population." *American Economic Review*, **101**(2), 664–94.

Wakker P, Erev I, Weber EU (1994). "Comonotonic Independence: The Critical Test Between Classical and Rank-Dependent Utility Theories." *Journal of Risk and Uncertainty*, **9**(3), 195–230.

Wallin A, Swait J, Marley AAJ (2017). "Not Just Noise: A Goal Pursuit Interpretation of Stochastic Choice." *Decision*.

Weiss LG, Saklofske DH, Coalson DL, Raiford SE (2010). *WAIS-IV Clinical Use and Interpretation: Scientist-Practitioner Perspectives*. Academic Press.

Wilcox NT (1993). "Lottery Choice: Incentives, Complexity and Decision Time." *The Economic Journal*, **103**(421), 1397–1417.

Wilcox NT (2008). "Stochastic Models for Binary Discrete Choice Under Risk: A Critical Primer and Econometric Comparison." In J Cox, GW Harrison (eds.), "Risk Aversion in Experiments," volume 12 of *Research in Experimental Economics*, pp. 197–292. Bingley, UK: Emerald.

Wilcox NT (2011). "Stochastically More Risk Averse: A Contextual Theory of Stochastic Discrete Choice Under Risk." *Journal of Econometrics*, **162**(1), 89–104.

Wilcox NT (2015a). "Error and Generalization in Discrete Choice Under Risk." *Working Paper*, Chapman University.

Wilcox NT (2015b). "Unusual Estimates of Probability Weighting Functions." *Working Paper*, Chapman University.

World Bank Group (2015). *World Development Report 2015: Mind, Society, and Behavior*. Washington, DC: World Bank.

Wright RA (1996). "Brehm's Theory of Motivation as a Model of Effort and Cardiovascular Response." In PM Gollwitzer and JA Bargh (eds.), *The Psychology of Action: Linking Cognition and Motivation to Behavior*, pp. 424–453. Guilford Press.

Wu G, Gonzalez R (1996). "Curvature of the Probability Weighting Function." *Management Science*, **42**(12), 1676–1690.

Vita

Aleksandr Alekseev was born on March 2, 1987 in Leningrad (now St. Petersburg), Russia. He studied Economics at St. Petersburg State University, where he earned his Specialist degree (BA equivalent, 5 year program), *summa cum laude*, in 2009. He continued studying Economics at the European University at St. Petersburg (Russia), where he earned his Master's degree in 2012. Aleksandr began pursuing a PhD in Economics in 2012, when he started a program at Emory University. In 2013, he transferred to Georgia State University (GSU) where he had the pleasure of working under the supervision of Dr. James Cox and Dr. Glenn Harrison.

Aleksandr's research primarily lies in the field of behavioral and experimental economics. Within this field, his interests include choice under risk, stochastic choice, labor economics and economics of human development, welfare economics, and methodology of experimental economics. Aleksandr has presented his work at the meetings of the Economic Science Association, Southern Economic Association, and Western Economic Association, as well as at research seminars at European University at St. Petersburg, Georgia State University, George Mason University, and University of Chicago. He has publications in such journals as *Journal of Economic Behavior and Organization*, *Annals of Finance*, and *Mathematical Social Sciences*.

At GSU, Aleksandr taught Principles of Microeconomics and Mathematics for Economists. Aleksandr is a recipient of an Outstanding Graduate Research Assistant Award, Center for the Economic Analysis of Risk Small Grant, Second Century Initiative University Doctoral Fellowship, Center for the Economic Analysis of Risk Scholarship, and Andrew Young School Dissertation Fellowship. Since 2015, he has been the co-founder of the Experimental Methods Forum, a student group at GSU dedicated to behavioral and experimental economics.

Aleksandr was awarded a PhD in Economics by Georgia State University in August 2018. He begins his work as a postdoctoral fellow at the Economic Science Institute at Chapman University starting September 2018.