EXPERIMENTS ON INEQUALITY: MEASURING INEQUITY AVERSION AND ITS IMPACT ON WELL-BEING

by

Lina Diaz
A Dissertation
Submitted to the
Graduate Faculty
of
George Mason University
In Partial fulfillment of
The Requirements for the Degree
of
Doctor of Philosophy
Economics

Committee:

_____ Director

_____

_____

_____

_____ Department Chairperson

_____ Program Director

_____ Dean, College of Humanities and Social Sciences

Date: _____ Spring Semester 2020
George Mason University
Fairfax, VA

Experiments on Inequality: Measuring Inequity Aversion and its Impact on Well-Being

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at George Mason University

By

Lina Diaz
Master of Science
Universidad de Los Andes, 2012
Bachelor of Science
Universidad de Los Andes, 2010

Director: Daniel Houser, Professor
Department of Economics

Spring Semester 2020
George Mason University
Fairfax, VA

Copyright © 2020 by Lina Diaz
All Rights Reserved

Dedication

I dedicate this dissertation to my loving family.

Acknowledgments

I would like to thank the Interdisciplinary Center for Economic Science (ICES) for providing a welcoming environment to learn and grow, especially its director Dr. Daniel Houser, who has believed in me, supported me throughout these years, and encouraged me with patience and optimism. I also thank other professors in my committee (in alphabetical order): Dr. César Martinelli, Dr. Kevin McCabe and Dr. Thomas Stratmann. I could have not asked for a better committee. I thank my coauthors Homa Zarghamee and John Ifcher, who taught me to do rigorous research through their example. Juan Camilo Cárdenas, Mariana Blanco and Luz Angela Rodríguez not only guided me in my first steps in research, they also ignited my love for experimental economics and they continue to be a source of inspiration.

Table of Contents

Page

List of Tables . . . vii
List of Figures . . . viii
Abstract . . . ix
1 First Chapter . . . 1
1.1 Introduction . . . 1
1.2 Literature Review . . . 5
1.2.1 Social Preference Models . . . 5
1.2.2 Fehr & Schmidt (1999) Model . . . 7
1.2.3 Previous Approaches to Estimating F&S Model . . . 7
1.2.4 Subjective Survey Data and Utility Measurement . . . 10
1.3 Methodology . . . 11
1.3.1 Inequity Aversion List . . . 11
1.3.2 Hypotheses . . . 13
1.4 Experimental Design . . . 14
1.4.1 First Part . . . 15
1.4.2 Second Part . . . 17
1.4.3 Questionnaire and Payment . . . 17
1.5 Empirical Analysis . . . 18
1.5.1 Parameter Estimation . . . 18
1.5.2 Results . . . 21
1.5.3 Demographic Characteristics . . . 26
1.5.4 Robustness Checks . . . 27
1.6 Discussion . . . 31
1.7 Conclusion . . . 32
2 Second Chapter . . . 34
2.1 Introduction . . . 34
2.2 Methodology . . . 35
2.2.1 Scoring Procedure . . . 38
2.3 Results . . . 38

2.4 Conclusion 40
3 Third chapter 41
3.1 Introduction 41
3.2 Literature Review 43
3.3 Experimental Design 46
3.3.1 Discovery and pilot studies 46
3.3.2 Online experiment 47
3.3.3 Own- and relative-rewards shocks 51
3.3.4 Experimental predictions 53
3.4 Results 56
3.4.1 Non-parametric tests of own- and relative-rewards shocks 57
3.4.2 Regression and subgroup analyses 58
3.5 Discussion 61
A Appendix A: Previous Approaches 66
B Appendix B: Tables 69
C Appendix C: Figures 80
D Appendix D: Monte Carlo Simulation 87
E Appendix E: Instructions 89
References 97
References 97

List of Tables

Table Page

2.1 Model of best fit for each participant 40

A.1 Instrument used by Blanco et al. (2011) 67

A.2 Instrument used by Yang et al. (2016) 68

B.1 Probit estimation (marginal effects) 69

B.2 Parameters' differences by demographic characteristics 70

B.3 Joint significance tests post Tobit estimation of equations (1.5) and (1.6) . 71

B.4 Scoring results by individual: Part 1 72

B.5 Scoring results by individual: Part 2 73

B.6 4X2 treatment-design, with own, Participant X's, and relative rewards by cell 74

B.7 Relative- and own-rewards shocks and formulae used for estimation 74

B.8 Demographic characteristics, n = 860 75

B.9 Pre- and post-revelation MSF-scores and ΔMSF by cell 76

B.10 Estimated impacts of relative- and own-rewards shocks on ΔMSF 77

B.11 Regression estimates of the impacts of relative- and own-rewards shocks on ΔMSF 78

B.12 Regression estimates of the impacts of relative- and own-rewards shocks on ΔMSF when including dropped subjects and when including pilot-study subjects, and on post-revelation 10-item BPNSS score 79

List of Figures

Figure		Page
1.1	Screenshot of the top of the *inequity aversion list*	12
1.2	Graphical representation of Fehr and Schmidt's model	18
1.3	Individual level results: α and β	20
1.4	Aggregate utility	22
1.5	Reversals (instances in which $\beta_i > \alpha_i$)	23
1.6	Parameters' distributions: List 1 vs. Games	24
1.7	Absolute change in List 1 responses	28
1.8	Parameters' distributions: Lists 1, 2 and 3	29
2.1	Possible behavioral models in the *inequity aversion list*	37
C.1	List 1	80
C.2	List 2	81
C.3	List 3	82
C.4	Aggregate utility List 2	83
C.5	Aggregate utility List 3	83
C.6	Weak adherents to Fehr and Schmidt's model	84
C.7	Non-adherents to Fehr and Schmidt's model	84
C.8	Individuals displaying maximin preferences	85
C.9	Individuals displaying efficiency concerns	85
C.10	Percentage change in List 1 responses	86
D.1	Distribution for 1 million tests	88

Abstract

EXPERIMENTS ON INEQUALITY: MEASURING INEQUITY AVERSION AND ITS IMPACT ON WELL-BEING

Lina Diaz, PhD

George Mason University, 2020

Dissertation Director: Daniel Houser

Inequality has been the focus of numerous studies in Economics and the subject of ongoing debate in the profession. However, natural environments often suffer from confounding effects and the mere difficulty of measuring wealth accurately. Thus, laboratory experiments constitute a valuable tool to examine effects difficult to study otherwise. This dissertation focuses on measuring the negative effect of inequality on individuals' well-being in the laboratory.

Chapter 1, proposes a novel methodology called the *inequity aversion list* to estimate the parameters of Fehr and Schmidt's (1999) model of inequity aversion. Previous elicitation methodologies often use standard games, like ultimatum and dictator, which results in an incongruity between estimates at the aggregate and individual levels. The explanation I offer for this discrepancy is that reciprocity and inequity aversion are confounded when analyzing other-regarding behavior in standard games, biasing the estimates. The *inequity aversion list* improves over previous methodologies by isolating the effect of inequity aversion from other motives. I take the *inequity aversion list* to the laboratory and find that subjects largely display the predictions of Fehr and Schmidt's model and behave according to the model's assumptions.

In chapter 2, I use the *inequity aversion list* to test various behavioral models beyond Fehr & Schmidt's (1999) inequity aversion. Using the data from the experiment explained in the first chapter, I identify the patterns of behavior individuals may display when answering the *inequity aversion list*. Each of these patterns is associated with a behavioral model. Using a scoring procedure, I determine which behavioral model fits each individual's data best.

Chapter 3, presents the results from an online experiment absent choice in measuring subjective well-being (SWB) before and after an exogenous shock that reveals to subjects how many experimental points they and another subject receive, and whether or not points are worth money. The result is that subjects are made significantly happier when they receive monetized rather than non-monetized points, suggesting money is valued more than the points it represents. In contrast, subjects are made equally unhappy when they receive fewer monetized points as when they receive fewer non-monetized points than others, suggesting relative money is not valued more than the relative points it represents.

Chapter 1: Eliciting Inequity Aversion[1]

1.1 Introduction

Social preferences have been largely studied as a motive driving behavior in environments as diverse as consumer response to price changes, attitudes toward different tax schemes, and employee response to changes in wages and employment practices (Charness & Rabin, 2002). One of the leading models devised to incorporate the analysis of other-regarding preferences is Fehr and Schmidt's (1999) inequity aversion model, where the authors extend the classical utility maximizing model to present a framework consistent with both purely selfish and other-regarding preferences.

The main attribute of Fehr and Schmidt (1999) model, hereafter F&S, is that it describes in a tractable and simple equation how fairness motives impact an individual's utility function. The model incorporates the distaste for both advantageous and disadvantageous inequity, captured by two separate parameters: α, which corresponds to disadvantageous inequity (*envy*); and β, which corresponds to advantageous inequity (*guilt*).

The envy and guilt parameters in F&S have been estimated using different methodologies and various samples across the world[2]. So far, the evidence collected seems puzzling: while the model captures the behavior of individuals at the aggregate level, it does not predict behavior at the individual level in a robust way.

Aggregate-level analyses are based on either the distribution of subjects' behavior across games or the estimation of structural models, capturing the behavior of an average subject while the heterogeneity is incorporated in the error (Engelmann & Strobel, 2004). However,

[1]This chapter is joint work with Dr. John Ifcher, Dr. Daniel Houser, and Dr. Homa Zarghamee.

[2]See Bellemare, Kröger, and Van Soest (2008, 2011) for a representative sample of Dutch adults. Laboratory studies with students include Goeree & Holt, 2000; Ellingsen & Johannesson, 2004; Daruvala, 2010 (This is not intended to be an exhaustive list).

Fehr and Schmidt's model is a model of an individual's utility, which may not necessarily describe behavior at the aggregate level. Furthermore, the individual heterogeneity in social preferences that has been observed in the laboratory calls for an individual-level analysis of fairness concerns. This point was highlighted by Andreoni and Miller (2002) as they observed individual level heterogeneity in a modified dictator game; in their experiment participants displayed behavior ranging from Utilitarian to Rawlsian to perfectly selfish.

Blanco, Engelmann, and Normann (2011) were the first study to implement a within-subject analysis to test the predictive power of Fehr and Schmidt (1999) at the individual level. Blanco et al. (2011) use responders' behavior in an ultimatum game to elicit the envy parameter and a modified version of the dictator game to elicit the guilt parameter[3]. They, then compare the elicited parameters with behavior in a public goods game and a sequential prisoner's dilemma, and find the aforementioned incongruity between individual and aggregate behavior.

As an alternative approach to games, Yang, Onderstal, and Schram (2016) propose the use of choice menus consisting of 10 decisions between 2 options. Each option constitutes a payoff for both the decision maker and an anonymous receiver. Although Yang et al. (2016) find support for F&S at the individual level, their parameters' estimates are affected by modifications to the range of values in the choice sets, and by changes in the relative income inequity (He & Wu, 2016).

In this paper, we propose a new methodology to estimate the parameters in F&S. Our methodology overcomes problems inherent to previous methodologies by using a different approach. We measure utility directly, in terms of expressed satisfaction, contrary to previous approaches based on revealed preferences. Our methodology is designed to withstand changes to either the range of options or the relative income inequity, as well as to isolate the effect of inequity aversion from other motives driving behavior in standard games.

[3]The public goods game (PG) has been used as well to estimate inequity aversion parameters (Ashley, Ball, & Eckel, 2010) but only at the aggregate level. The only study we are aware of that estimates parameters at the individual level with a PG is Tasch and Houser (2018).

Let us explain what we mean by other motives. Fairness concerns can be of two different types: *outcome-based* and *procedure-based*. On one hand, an individual influenced by outcome-based fairness concerns has a distaste for unequal payments, or any other outcome, regardless of the specific steps leading up to those payments. On the other hand, an individual influenced by procedure-based fairness concerns cares about the context and the resource allocation mechanism, thus the process itself and the intentions held by decision makers matter. In short, inequity aversion can be seen as an instance of outcome-based fairness concerns, while reciprocity can be seen as an instance of procedure-based fairness concerns. Either or both types of fairness concerns can affect decisions in the games commonly used to estimate F&S parameters: ultimatum and dictator.

Our methodology focuses on outcome-based fairness concerns by estimating the parameters in F&S as measures of inequity aversion only. Ultimately, F&S is a model of outcome-based fairness concerns, where the model parameters accompany an average of the absolute difference between payments obtained by an individual and her counterparts. Our methodology uses the simplified two-player version of F&S model to estimate the parameters, based on the payment difference between the two players.

We ask participants to evaluate their satisfaction level with 21 different allocations between themselves and a person they are randomly paired with. Participants perform these evaluations using a slider for each allocation on a Likert scale. Then, we estimate envy (guilt) as the line of best fit's slope for those allocations where the individual experiences disadvantageous (advantageous) inequity. A thorough explanation of our methodology is presented in the experimental design section.

To analyze our results we focus on two aspects: first, we check the robustness of our methodology; second, we analyze our participants' degree of consistency with F&S. The first part of the analysis is crucial as we are proposing a new methodology, thus we want to verify whether it captures inequity aversion in a consistent way. The second part of the analysis addresses the incongruity between individual and aggregate behavior, by checking consistency with the model and its assumptions for each individual and for the aggregate.

Regarding the first part of the analysis, we find that changes to the information shown to participants, in terms of the content and the way information is presented, don't change our results. As for the second part of the analysis, we find consistency with F&S at both the aggregate and the individual levels. In the empirical analysis section we propose two definitions of consistency with F&S in the context of our methodology: a weak version and a strong one. Our results indicate that 75% (65%) of participants in our sample display preferences weakly (strongly) consistent with F&S. Also, compared with previous elicitation methods, our methodology has greater consistency with one of the model's assumptions, i.e. that the envy parameter is larger or equal to the guilt parameter. This assumption is also met at the aggregate level.

This paper's contribution to the current literature is both methodological and empirical. We propose a novel methodology to estimate the inequity aversion parameters in F&S. Our methodology improves upon previous methodologies by solving the apparent incongruity between the aggregate and the individual levels. We hereby offer a tool that isolates the effect of inequity aversion from procedure-based fairness concerns. It is our hope that this tool benefits experimental economists who want to collect data on subjects' inequity aversion for any experiment where fairness concerns influence individual decision making.

The rest of the paper is organized as follows. In section 2, we discuss the literature that our study speaks to and explain how our paper relates to this literature. In section 3, we describe our methodology called *inequity aversion list* and present the hypotheses to be tested. In section 4, we describe the experimental design. In section 5, we present the empirical analysis, explaining further the parameter elicitation methodology and discussing the results we obtain. Finally, in sections 6 and 7, we conclude and discuss further research.

1.2 Literature Review

1.2.1 Social Preference Models

Understanding social preferences is important not only in terms of making sense of findings inside the laboratory, but also for understanding human behavior more broadly. Insights from other-regarding preferences explain behavior of diverse economic phenomena as: "...the persistence of non-competitive wage premiums, the incompleteness of contracts and the absence of explicit incentive schemes, the allocation of property rights, the conditions for successful collective action, and the optimal design of institutions." (Fehr & Schmidt, 2006, p. 618).

Social preference models focus on different pro-social attitudes an individual may hold. These include inequity aversion (Loewenstein, Thompson, & Bazerman, 1989; Fehr & Schmidt, 1999; Bolton & Ockenfels, 2000), reciprocity and the role of others' intentions (Rabin, 1993; Charness & Rabin, 2002; Dufwenberg & Kirchsteiger, 2004; Falk & Fischbacher, 2006; Segal and Sobel, 2007, 2008), efficiency[4](Charness & Rabin, 2000), and the type of person an opponent is (i.e to what degree the other person is altruistic or spiteful) (Levine, 1998).

Several studies have compared the relative explanatory power of these models in various environments inside the laboratory. However, no single model is able to encompass all the behavior observed, although for different samples different social preference models fit the data better than others (See Engelmann & Strobel, 2004; Daruvala, 2010).

Among the social preference models, F&S stands out as being simple and tractable, yet consistent with a wide range of behavior observed in competitive settings and standard games. Those games include ultimatum, gift exchange, trust, and public good games (Fehr & Schmidt, 2006). Nonetheless, F&S describes behavior that extends beyond the laboratory. The most natural application is that of peer inequity aversion in the workplace. In

[4]surplus maximization

settings where workers are able to acquire information about their peer salary[5], earning less has a negative effect on effort (Cullen & Perez-Truglia, 2018; Cohn, Fehr, Herrmann, & Schneider, 2014), output (Cullen & Perez-Truglia, 2018; Breza, Kaur, & Shamdasani, 2018), attendance (Breza et al., 2018), retention (Cullen & Perez-Truglia, 2018; Dube et al., 2019), and job satisfaction (Card, Mas, Moretti, & Saez, 2012). Interestingly, Breza et al. (2018) find that those negative effects vanish when productivity differences are large and observable (justified inequality).

Recent literature on social preferences has focused on understanding how social preferences emerge, and whether they are context dependent. Sutter, Zoller, and Glätzle-Rützler (2019) provide a survey of experiments with children and adolescents; the overall behavior in ultimatum and trust games shows that fairness concerns are developed early, although in the dictator game inequity aversion preferences are displayed only until later childhood. In particular, Fehr, Bernhard, and Rockenbach (2008) find that inequity aversion is more frequent from ages 3 to 8 compared to spiteful and altruistic preferences. Furthermore, Blake et al. (2015) compare children's behavior in seven diverse societies and find that while disadvantageous inequity aversion emerges across all populations by middle childhood, advantageous inequity aversion only emerges in three populations, and only later in development. In terms of how much social preferences change depending on the context, Hedegaard, Kerschbamer, Müller, and Tyran (2019) find stability of preferences across games and over time.

Other studies have focused on classifying subjects into social-preferences types. Kerschbamer's (2015) equality equivalence test (EET) sorts people into nine archetypes of distributional concerns[6]. Similarly, the social value orientation (SVO) classifies people on a continuum going from altruistic to competitive. See Murphy, Ackermann, and Handgraaf (2011) for a measure of SVO on a continuous scale.

[5]Studies in this topic include field experiments where part of the intervention consists precisely in providing this information to workers. Dube, Giuliano, and Leonard (2019) use quasi-experimental variation from a rule-based formula for pay raises.

[6]The archetypes are spiteful, kick-down, equality averse, envious, selfish, kiss-up, inequity averse, maximin and altruistic.

1.2.2 Fehr & Schmidt (1999) Model

F&S describe individual i's utility function, in an environment of n players, as follows.

$$U_i(x) = x_i - \alpha_i \frac{1}{n-1} \sum_{j \neq i} \max\{x_j - x_i, 0\} - \beta_i \frac{1}{n-1} \sum_{j \neq i} \max\{x_i - x_j, 0\} \tag{1.1}$$

Where x_i is individual i's monetary payoff and x_j is individual j's monetary payoff for $i \neq j$. This model has two parameters that capture inequity aversion depending on the direction of the payoff difference: disadvantageous inequity is captured by α_i, which can be interpreted as i's envy level; while advantageous inequity is captured by β_i, which can be interpreted as i's guilt level. Evidently, in the absence of payoff differences between i and j utility is entirely determined by an individual's own payoff.

The model posits the following assumptions over the parameters: $\alpha_i \geq \beta_i \geq 0$ and $\beta_i < 1$. The first assumption implies that subjects suffer more from disadvantageous inequity than from advantageous inequity while they don't enjoy either type of inequity. The second assumption indicates that individual i puts more weight on her own payoff x_i than the other player's payoff x_j.

1.2.3 Previous Approaches to Estimating F&S Model

Our approach is closest to a set of experiments focused on the estimation of the F&S parameters at the individual level, in particular Blanco et al. (2011) and Yang et al. (2016) (See appendix A for a description of these approaches.) However, the envy and guilt parameters in F&S have been typically estimated at the aggregate level, i.e. capturing estimates of the average for a sample of participants or an aggregate distribution. The reason for this type of analysis is the use of either analysis of the sample distribution (Fehr & Schmidt, 1999; Ellingsen & Johannesson, 2004) or a structural model framework (Bellemare, Kröger, & van Soest 2008, 2011; Goeree & Holt, 2000). In both cases, the analysis is based on behavior observed in various games and the average coefficient for envy (guilt) varies from

0.31 to 1.89 (0.34 to 0.80)[7]. Additionally, these aggregate-level analyses show that the envy parameter is typically larger than the guilt parameter, in line with F&S's assumption.

A common practice in studies eliciting F&S parameters is the use of standard games, such as ultimatum and dictator. In this paper we argue that using standard games to estimate F&S parameters may result in confounding inequity aversion with other motives.

In the standard ultimatum game (UG), the decision to reject not only entails an inequity aversion concern, it also serves as a mechanism to (i) punish the Proposer and (ii) express negative emotions. The first mechanism is related with the notion of *altruistic punishment*: the Responder punishes a violation of a social fairness norm at the expense of his own pecuniary return (Fehr & Fischbacher, 2003). The second mechanism refers to *negative emotions* in reaction to low offers in the UG. Low offers elicit activity in brain areas associated with negative emotional states (Sanfey, Rilling, Aronson, Nystrom, & Cohen, 2003)[8].

Modifications to the standard UG have provided supporting evidence to both *altruistic punishment* and *negative emotions* as explanations for rejections in UG. When offers are effectuated by a random generator device, low offers are accepted more frequently than in the standard UG (Blount, 1995; Bellemare et al., 2011). Similarly, when Responders are allowed to express their negative emotions through a message (Xiao & Houser, 2005) or are given a cooling-off period before making a decision (Grimm & Mengel, 2011) low offers are also accepted more frequently than in the standard UG.

In a modification of the ultimatum game known as the cardinal ultimatum game (Bolton & Zwick, 1995), or the ultimatum minigame (Gale, Binmore, & Samuelson, 1995), senders' choices are reduced to only two possible allocations while receivers maintain the same choice: deciding whether to accept or reject. In this environment Falk, Fehr, and Fischbacher (2003) find that identical offers are rejected more frequently when the proposer has more fair alternatives available to her, than if only more unfair alternatives are available. Brandts

[7]Eckel and Gintis (2010) consolidate results from several experiments estimating F&S parameters at the aggregate level.

[8]Specially the bilateral anterior insula, shown to be involved in the experience of anger and disgust (Sanfey et al., 2003).

and Solà (2001), and Güth, Huck, and Müller (2001) also find supporting evidence for the role of intentions using the same type of game.

These experiments highlight the importance of perceived intentions on Responders' behavior. A Responder concerned with the Sender's intentions may reject an offer if she considers it unfair, displaying negative reciprocity by matching a deviation of a social norm with reciprocal behavior. This behavior is indistinguishable from the same rejection triggered by distributional concerns of an inequity averse Responder. Therefore, the ultimatum game may overestimate the envy parameter in F&S by combining effects that are pointing in the same direction: *inequity aversion* and *negative reciprocity*[9].

Similarly, the dictator game may overestimate the guilt parameter as the dictator meets expectations of reciprocity when adhering to a certain social norm. Hence, dictators send smaller amounts to their counterparts when they have to earn their endowment (Cherry, Frykblom, & Shogren, 2002), and when a double blind procedure is implemented (Hoffman, McCabe, & Smith, 1996). Even though there is no direct reciprocity between the dictator and her counterpart, these modifications to the standard dictator game are evidence of behavior consistent with procedure-based fairness concerns.

Another possible confound that may bias parameter estimates in F&S is the *own-income effect*. Players in ultimatum and dictator games make decisions that affect the payment difference, but those decisions also impact their own income. This effect would potentially shrink the parameters in F&S by transferring weight from fairness concerns to self-interest. Our methodology is shielded against this potential confound as the decision maker's payoff is held constant[10].

Our approach contributes to the current literature by providing an instrument that improves upon previous elicitation methods in two important ways. First, we isolate the

[9] Although negative reciprocity is strongly affected by perceived intentions, evidence from simulated labor markets, social dilemma games and trust games finds weak support for intentions-driven *positive* reciprocity (Falk et al., 2003)

[10] In an experiment comparing the relative importance of F&S, Bolton and Ockenfels (2000), efficiency concerns and maximin preferences, Engelmann and Strobel (2004) find that the relative importance of the different motives does not change in the case where subjects make choices over distributions keeping their own payoff fixed, from the case where subjects make choices over distributions with different payoffs for themselves.

effect of inequity aversion from other motives related with procedure-based fairness concerns. Second, our measure is robust to modifications of the values presented to participants.

1.2.4 Subjective Survey Data and Utility Measurement

We call our instrument the *inequity aversion list*. It is designed to estimate the parameters in F&S based on the utility individuals attain from monetary allocations with different levels of inequality. We use satisfaction as a proxy for utility in our empirical strategy.

The shortcomings of reported subjective data are well recognized by economists: susceptibility to the order of the questions, the wording of the questions, or the scales applied should certainly be acknowledged (See Krueger and Schkade (2008), Duckworth and Yeager (2015) and Brañas-Garza, Galizzi, and Nieboer (2018) for discussions on this topic). Recognizing these biases, however, does not impede the usage of satisfaction measures. On the contrary, it strengthens it by allowing a better use of subjective survey data, as proposed by Bertrand and Mullainathan (2001) and discussed by Frey and Stutzer (2002).

A more fundamental question perhaps is whether individuals exert mental effort towards answering these type of questions: Do they read the instructions thoroughly and respond truthfully? This is an empirical question; in fact, our experimental design addresses this question as explained in Section 1.4 *Experimental design*. Moreover, wariness of satisfaction measures should not be greater than other measures used by economists; Charness and Grosskopf (2001) make the point that mainstream economics analysis has typically used self-reported data from unemployment and Census surveys[11].

This study also speaks to a growing literature on happiness from the perspective of Economics. In this literature, Subjective Well-being (SWB) is typically used as a proxy for utility. Benjamin, Heffetz, Kimball, & Rees-Jones (2012, 2014) and Benjamin, Heffetz, Kimball, and Szembrot (2014) find evidence indicating that, although SWB is an important

[11]Examples of Economics research utilizing questions in which people ascertain their general level of satisfaction, or a domain specific level of satisfaction, abound in Economics. Domain specific questions are applied to topics as varied as health, financial situation, housing, marriage and job (Van Praag & Ferrer-i-Carbonell, 2004). For example, self reported measures of job satisfaction and job-related subjective well-being are used as predictors of labor mobility (Green, 2010).

argument in the utility function, individuals do not maximize SWB exclusively. Thus refining SWB measurement is important, specially given that SWB is multidimensional and it may have a different weight on utility depending on the type of decision an individual is making[12]. Our *inequity aversion list* measures SWB taking into account negative affect (in the domains of envy and guilt) and satisfaction. This study contributes to the happiness literature taking a step further in the direction of refining the measure of SWB in the particular setting of experimental payoffs.

1.3 Methodology

1.3.1 Inequity Aversion List

The *inequity aversion list* is a tool designed for laboratory experiments. It displays 21 allocations between a participant i and an anonymous person in the room j. The payoff for i is fixed at \$20 for every allocation in the list, while the payoff for j goes from \$10 to \$30. Every participant is asked to complete the list in the role of i, indicating her satisfaction level for each allocation between i and j on a scale from 1=*extremely dissatisfied* to 7=*extremely satisfied*. Any number between 1 and 7 can be chosen, up to one decimal place, by moving the slider with the mouse. Numbers to the right of each slider display the current position of the slider to assist participants in rating their satisfaction accurately. Sliders can be adjusted an unlimited number of times and the responses to the 21 allocations can be given in any order. There is an 'okay' button at the bottom of the screen. Once participants click that button, they cannot change their responses; only final responses are recorded. Participants are required to provide an answer for each allocation before they are allowed to proceed.

Figure 1.1 presents a screenshot of the first 10 allocations in the *inequity aversion list* as shown to participants in the laboratory (See figure C.1 in the Appendix for the entire list). Each slider moves independently and responses are not required to be monotonic.

[12]SWB is a better approximation of utility in minor decisions, like the ones we present in this paper, than in important life decisions (Benjamin et al., 2012).

Thus the computer program allows a response where the fourth allocation yields a lower satisfaction level than the third and the fifth, as shown in figure 1.1 for illustrative purposes. We emphasize this point to clarify that we are not enforcing consistency with F&S model by restricting participants answers in any way. Furthermore, we decided to present allocations in the list maintaining an increasing order over j's payoff to reduce the cognitive burden on participants. This does not obviously imply that participants are consistent with F&S as the sliders function independently from each other, and we emphasize in the instructions that there are no wrong answers.

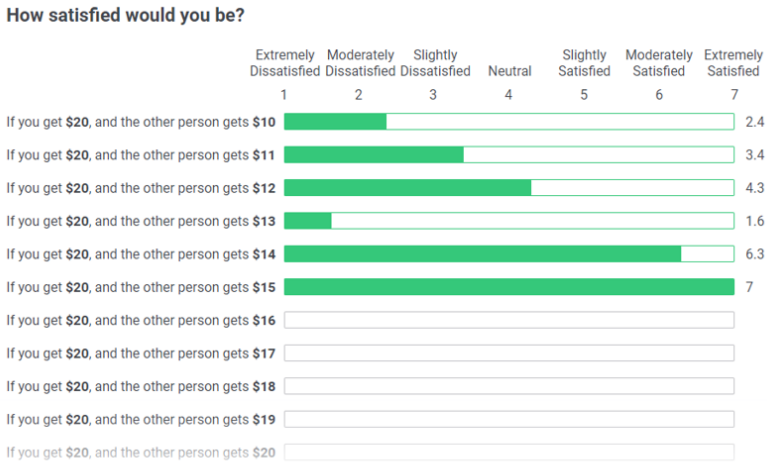


Figure 1.1: Screenshot of the top of the *inequity aversion list*

Note: In the experiment participants can see the entirety of the list by scrolling down with their mouse.

When participants evaluate the allocations in the *inequity aversion list*, they all do it from the perspective of i. However, to determine the payments, not all participants can be i. In each pair, one participant is randomly selected to be i and the other is randomly selected to be j. Then, one of the allocations in the list is randomly selected to be implemented for the pair. This means that participant i receives \$20, and participant j receives the corresponding value of the randomly selected allocation in the list, e.g. if the selected

allocation is the first one in the list, j receives \$10.

It warrants mention that the responses from the list do not affect the probability of selecting certain allocation for payment. In this way, we rule out strategic responses misrepresenting true preferences in order to increase the probability of certain outcome. Additionally, the instructions participants see do not include the labels i and j that we are using here. From the participants' perspective, they report their satisfaction levels with each allocation in the *inequity aversion list*, and at the end of the experiment two randomizations determine their payment (See appendix E for instructions given to participants in the experiment.)

The list includes one allocation in which both participants get the same payoff. The rest of the allocations are evenly split into allocations where i gets more than j and allocations where j gets more than i. This allows us to elicit the guilt parameter from the first half of the allocations and the envy parameter from the second half.

Our *inequity aversion list* has several advantages. First, it isolates the effect of inequity aversion from alternative other-regarding motives. Secondly, as it will be shown, it is robust to modifications in the list's values. Thirdly, this methodology allows us to find point estimates of the parameters instead of upper and lower bounds. Finally, as the task is easy to explain and understand, it takes a short time to complete and is easy to implement in any experiment as a complementary measure of an individual's preferences.

1.3.2 Hypotheses

The *inequity aversion list* allows us to test individually whether participants in our experiment display inequity aversion as described by F&S. Furthermore, we can test the effect of individual characteristics on inequity aversion. Fehr et al. (2006) find that women and non-economists choose egalitarian allocations more often than men and economists.

- Hypothesis 1: Individuals behave consistently with F&S model.
  - i Individuals maximize their utility at equality.
  - ii Individuals are more sensitive to envy than guilt, i.e. $\alpha_i \geq \beta_i$

- Hypothesis 2: Certain individual characteristics have an effect on inequity aversion
    - i Women behave consistently with F&S model more often than men.
    - ii Non-economists behave consistently with F&S model more often than economists.

1.4 Experimental Design

We ran seven, 90 minutes long, experimental sessions at the ICES (Interdisciplinary Center for Economic Science) laboratory in George Mason University, all conducted on weekdays between September and October 2018. The experiment was programmed in Qualtrics. All decisions were anonymous and participants were paid privately and in cash. On average, participants received $22.90, including a payment of $5 for showing up.

A total of 106 students participated in the experiment, 6 of which were graduate students and the remainder were undergraduate students. Participants were recruited by email from a preregistered pool of students willing to receive invitations to attend experiments at George Mason's experimental economics laboratory. We allowed students of all majors and all class years to participate in the experiment.

The experiment consisted of two parts and a demographic questionnaire at the end. In the first part, participants completed 5 independent tasks. In each task, participants were randomly matched with another participant in the room. They were matched with a different person for each task and their identities remained anonymous. In the second part, participants repeated one of the tasks from the first part with a slight modification. In the following subsections, we describe in detail each part of the experiment.

Participants received general instructions at the beginning of the experiment highlighting the anonymity of their decisions, giving a broad explanation about the payment procedure, and some general rules were explained. They were also instructed on how to interact with the Qualtrics interface[13]. Specific instructions were provided as the experiment proceeded.

[13]Qualtrics provides an intuitive interface that doesn't require explanation for anyone having a basic knowledge of computers. Thus, we simply warned participants that after clicking an 'okay' button to proceed to the next task they wouldn't be able to go back to the previous task. We also warned them that they may have to scroll down in some cases to complete a task and see the 'okay' button to proceed.

Participants kept a physical copy of the general instructions throughout the experiment, and read the specific instructions for each task on their screens. To guarantee that they read and understood the instructions, a quiz was administered after some of the tasks to measure their level of understanding. The quiz was graded by the computer and participants were not able to proceed before passing the quiz. Participants were given three attempts at answering the quiz on their own. If participants failed after three attempts, the system locked by asking for a password that only experimenters knew; this allowed experimenters to visit participants' computer stations and clarify any questions participants may had. Experimenters made sure participants' questions were clarified before unlocking their screens by typing the password[14].

1.4.1 First Part

There were two categories of tasks in the first part of the experiment: lists and games. The category's order was randomized at the individual level while the order of the tasks within the categories was kept equal for all participants. The tasks were randomized at the category level and not at the task level in order to facilitate instructions' communication by keeping similar tasks together. Therefore, there were some general instructions for the lists and some general instructions for the games. 54 participants completed the lists before the games, and 52 completed the games before the lists; we found no significant difference between these two groups.

The first category consisted of 3 lists (see Appendix C for screenshots of each one of the lists), with the following range of payoffs for x_i and x_j:

- List 1: $x_i = \$20$, $x_j \in [\$10, \$30]$
- List 2: $x_i = \$10$, $x_j \in [\$0, \$20]$
- List 3: $x_i = \$20$, $x_j \in [\$1, \$39]$

List 1, was the *inequity aversion list* presented in the previous section. Lists 2 and

[14] 70.7% participants in our sample needed their screens to be unlocked in either the ultimatum or the modified dictator game.

3 were modifications to List 1 in the spirit of He and Wu (2016). List 2 consisted of a change in the relative income inequality by subtracting \$10 from all payoffs in List 1, i.e. a reduction for both individuals (i and j). F&S accounts for absolute inequality but not for relative inequality. Therefore, the model predicts that if absolute inequality is maintained, a change in relative inequality should not alter the parameters. List 3 extended List 1 to include 9 additional allocations at the beginning of the list and 9 at the end. If our inequity aversion list is a good instrument to elicit the parameters of the model, the inclusion of additional allocations should not alter the measures elicited with the original list.

The second category of tasks were two games: the ultimatum game and a modified dictator game. Both games were played using the strategy method. We followed closely the procedure by Blanco et al. (2011), described in appendix A, except for one difference: they allowed multiple switching points while we only allowed one. We decided to enforce a single switching point to facilitate data analysis[15].

In the ultimatum game, participants made two decisions: how much to send as the first player, and the minimum amount they would be willing to accept as the second player. In the modified dictator game, participants chose between increasing levels of equality, going from \$0-\$0 to \$20-\$20, and the alternative of keeping the \$20 for themselves while leaving the other person with \$0. The decision participants made as dictators was the minimum equal amount they were willing to accept instead of keeping \$20 for themselves. (See Appendix E for the instructions administered in both games.)

The switching point chosen by the receiver in the ultimatum game corresponds to a range of values the envy parameter lies in, and similarly the switching point chosen in the modified dictator game corresponds to a range of values the guilt parameter lies in. Therefore, in order to elicit the parameters' ranges individuals must display transitivity, i.e. a single

[15]While there is no consensus yet as whether single switching points should be enforced in experiments, Nielsen and Rehbeck (2019) present suggestive evidence of the advantage of single switching point enforcement. In their experiment, they measure direct preferences over transitivity, and other fundamental axioms. When revealed preferences conflict with stated axiom preferences, they offer participants the possibility of revising their choices over lotteries to make them consistent with the axioms. They find that 80% of subjects who show intransitivities in the form of multiple switching points change their choices to be transitive. Nielsen and Rehbeck conclude that enforcing a single switching point might actually help subjects express their underlying desire for transitivity.

switching point[16]. In contrast, the *inequity aversion list* relaxes the transitivity requirement by combining answers to several allocations in the list to elicit the point estimates for each parameter.

1.4.2 Second Part

To test whether answers to the *inequity aversion list* in the first part of the experiment were random, participants were asked to replicate their answers to List 1 ($x_i = \$20$, $x_j \in [\$10, \$30]$) in the second part of the experiment. Random answers would likely be more difficult to replicate than answers revealing true preferences. The payment for this task was based on accuracy: for each allocation, participants received \$0.25 if their answer in the second part was within 0.5 units away from their answer in the first part. Note that participants received instructions for the second part of the experiment only after completing the first part. Therefore, behavior in the first part of the experiment could not possibly be modified in anticipation of the task in the second part of the experiment.

1.4.3 Questionnaire and Payment

To conclude the experiment, participants completed a questionnaire and were paid. The questionnaire included demographic characteristics such as gender, age, major, minor, GPA, political views, parents' education, and income. After completing the questionnaire, participants received a \$5 show up fee and a payoff corresponding to each part of the experiment. For the first part, one task was randomly selected for payment. For the second part, participants received the sum of all List 1's allocations in which they replicated their answers given in part 1 of the experiment.

[16] In Blanco et al. (2011) 11 subjects, out of 72, displayed intransitive preferences and were ultimately dropped from the analysis.

1.5 Empirical Analysis

Unless stated otherwise, all the results presented in this section use data from List 1. Lists were presented in ascending order to all individuals. Therefore, all participants saw List 1 first, and thus we consider List 1 the cleanest measure of participants' preferences.

1.5.1 Parameter Estimation

The two-player version of F&S inequity aversion model is characterized by the following equation.

$$U_i(x) = x_i - \alpha_i \max\{x_j - x_i, 0\} - \beta_i \max\{x_i - x_j, 0\} \tag{1.2}$$

This equation can be represented graphically as Figure 1.2 shows, where the slopes of the two segments in the graph are the point estimates for α_i and β_i. The data points before equality are used to estimate the guilt parameter β_i, and those after equality to estimate the envy parameter α_i.

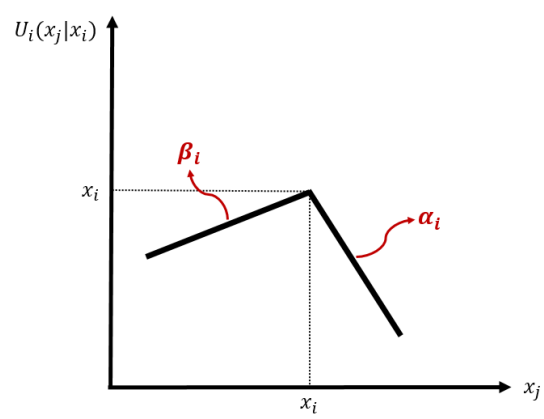


Figure 1.2: Graphical representation of Fehr and Schmidt's model

The responses from our *inequity aversion list* allow us to construct Figure 1.2 for each individual in our sample. Using the data points where $x_i \geq x_j$, we calculate the line of best

fit that corresponds to the first segment in figure 1.2. Likewise, we calculate the second segment based on the data points where $x_j \geq x_i$. Each segment's slope constitutes one of the model's parameters: guilt for the first segment and envy for the second one.

In practical terms, we estimate equations (1.3) and (1.4) for each participant i using an OLS model, where k indicates the allocation in the list. $\bar{x}$ is individual i's own payoff, it is fixed at the same level for all allocations k in the list and for all subjects. Hence, the first term in both (1.3) and (1.4) is a constant. x_{jk} denotes the other player's payoff for each allocation k. D_{1k} and D_{2k} are dummy variables that help identify separately the slopes of advantageous and disadvantageous inequity. $D_{1k}x_{jk}$ and $D_{2k}x_{jk}$ are interaction terms of the first two variables in each equation.

$$Satisfaction_{ik} = \gamma_{0i}\bar{x} + \gamma_{1i}x_{jk} + \gamma_{2i}D_{1k} + \gamma_{3i}D_{1k}x_{jk} + u_k \tag{1.3}$$

$$Satisfaction_{ik} = \delta_{0i}\bar{x} + \delta_{1i}x_{jk} + \delta_{2i}D_{2k} + \delta_{3i}D_{2k}x_{jk} + u_k \tag{1.4}$$

$$D_{1k} = \begin{cases} 1, & \text{if } x_j \leq \bar{x} \\ 0, & \text{otherwise} \end{cases}$$

$$D_{2k} = \begin{cases} 1, & \text{if } x_j \geq \bar{x} \\ 0, & \text{otherwise} \end{cases}$$

Note that estimates in equations (1.3) and (1.4) have a subscript i. The reason is that these equations are estimated for each individual separately. Therefore, the estimation process consists in using an individual's data to estimate β_i from (1.3), α_i from (1.4), and then repeat the process for each individual.

F&S parameters are calculated as the marginal effects of x_{jk} over $Satisfaction_{ik}$. For example, β_i is estimated over allocations where $x_j \leq \bar{x}$, and thus $D_{1k} = 1$. By substituting $D_{1k} = 1$ in (1.3) we obtain the marginal effect of x_{jk} over $Satisfaction_{ik}$ as $\beta_i = \gamma_{1i} + \gamma_{3i}$.

The corresponding exercise for (1.4) results in $\alpha_i = \delta_{1i} + \delta_{3i}$. Note that this methodology allows us to find point estimates, unlike previous studies estimating parameters' bounds, e.g. Blanco et al. (2011) and Yang et al. (2016).

Figure 1.3 summarizes the results for subjects who are consistent with F&S. We introduce our definition of consistency with F&S in the next section. For the rest of the subjects, we would need additional assumptions in order to calculate α and β, thus we simply abstain from making arbitrary assumptions. Nonetheless, in the next section we also discuss the behavior of those subjects who are not consistent with F&S.

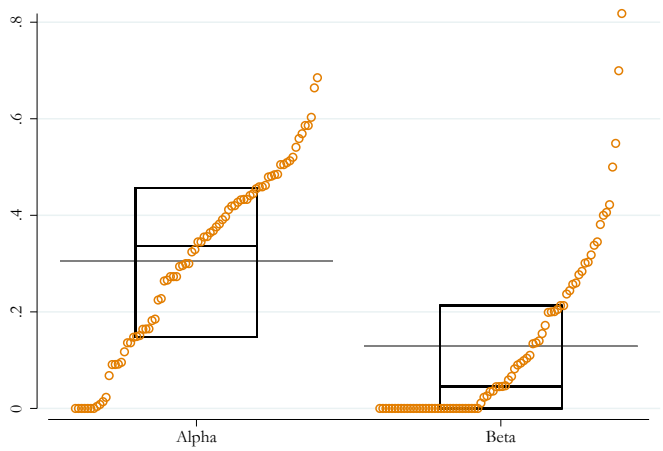


Figure 1.3: Individual level results: α and β

Note: This graph presents results for subjects consistent with F&S (N=80). Black boxes indicate the upper and lower quartiles, as well as the median. Horizontal gray lines, longer than the boxes, correspond to the mean. Orange dots display quantile plots. Alpha and beta are statistically different from zero (Wilcoxon matched-pairs signed-rank tests: $Z = 7.706$, $p = 0.000$; $Z = 6.669$, $p = 0.000$), and alpha is larger than beta (Sign test of matched pairs: $p = 0.000$).

1.5.2 Results

Consistency with F&S

The first step in our analysis consists of determining whether the participants in our sample display the behavior predicted by the model. Before presenting the results, let us first introduce our definitions of consistency with the F&S model.

> *Weak adherence to F&S model*: An individual weakly adheres to F&S model if her maximum level of utility is reached at equality. This is in spite of the level of utility, and whether the maximum utility is also reached at unequal allocations.

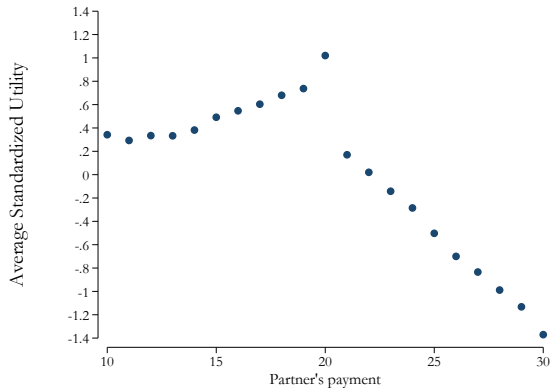
> *Strong adherence to F&S model*: An individual strongly adheres to F&S model if she displays weak adherence, and the model's assumption with regard to the relationship between the parameters is met, i.e. $\alpha_i \geq \beta_i$.

Our main result is that we find large consistency with F&S model, i.e. we cannot reject hypothesis 1. On the aggregate, we find strong adherence to F&S model: participants in our sample maximize their utility at equality and display more sensitivity to envy than guilt, i.e. $\alpha \geq \beta$. Figure 1.4 depicts the average level of utility for each allocation, where individual i's payment is fixed and individual j's payment goes from \$10 to \$30. As the figure shows, utility of disadvantageous inequity (to the right of \$20) is consistently lower than advantageous inequity (to the left of \$20). Furthermore, there is a drop in utility in the close vicinity of equality, yet the drop is substantially larger in the direction of disadvantageous inequity than in the opposite direction.

At the individual level, we also find consistency with F&S model. 75.5% of participants in our sample (80 participants out of 106) weakly adhere to F&S model[17], while 65% (69 participants) strongly adhere to F&S model[18].

[17] We show individual graphs for the group of weakly consistent participants in figure C.6 on the appendix.

[18] This figure is very similar even if the definition of strong adherence becomes stricter including all restrictions F&S imposed over the parameters. In other words, imposing $\alpha_i \geq 0$ and $1 > \beta_i \geq 0$ besides $\alpha_i \geq \beta_i$ doesn't change the fact that the majority of participants in our sample strongly adhere to F&S model. To be more precise, only one participant meets the definition of strong adherence we offer but wouldn't meet a stricter version.



N=106, $\alpha = 0.20$, $\beta = 0.06$

Figure 1.4: Aggregate utility

Note: Standardized utility is measured as $\frac{(U_{ik}-\overline{U_i})}{sd(U_i)}$, where U_{ik} is the utility expressed by individual i with regard to allocation k.

Reversals

There are 11 participants in our sample that are weakly adherent to F&S but display larger sensitivity to guilt than envy, as opposed to F&S's assumption that people are loss averse in social comparisons. We call violations of this assumption *reversals*.

This brings us to our second result: we find fewer reversals with our methodology than previous studies using standard games (Blanco et al., 2011) and choice menus (Yang et al., 2016). See figure 1.5. Finding fewer reversals is important insofar as it reflects greater consistency with the model. Moreover, the assumption that $\alpha_i \geq \beta_i$ is not only sensible, it was proposed by F&S because it is consistent with empirical evidence.

Inequity Aversion List versus Standard Games

Consistency with F&S model, and its assumptions, stands in contrast with findings from previous studies using standard games (Blanco et al., 2011). A comparison of the parameters captured by our methodology with those captured by standard games can shed light on

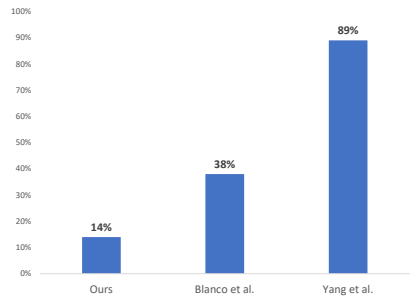


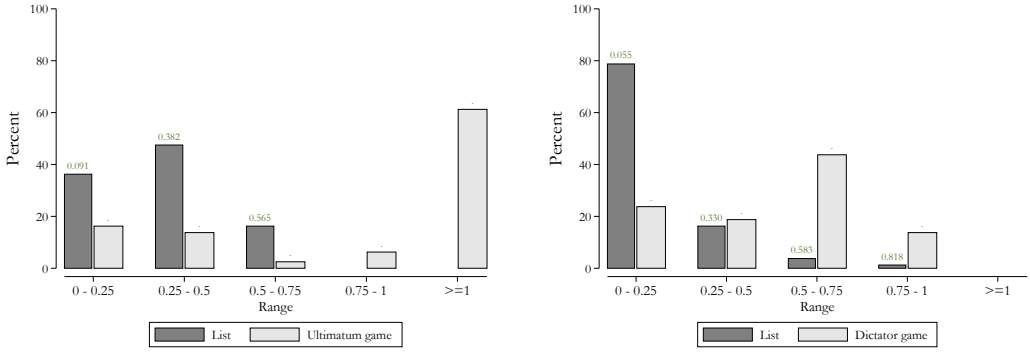
Figure 1.5: Reversals (instances in which $\beta_i > \alpha_i$)

the basis of the discrepancy. Figure 1.6 presents this comparison for each parameter. In both cases, using games for the elicitation results in larger parameter values. Note that these are distributions for the same group of individuals, who were asked to complete the *inequity aversion list* as well as the ultimatum and dictator games in the first part of the experiment[19].

We interpret the differences in results between standard games and the *inequity aversion list* as differences in what the estimates are capturing. In both cases, the parameters capture other-regarding preferences. However, the *inequity aversion list* only captures inequity aversion, while ultimatum and dictator games may be capturing simultaneously inequity aversion and some type of reciprocity. Fehr and Schmidt (1999) discuss this matter when they present the model: they indicate that their model does not distinguish the source of the fairness concern, thus α and β can be interpreted as a concern for equality or as a reduced-form concern for intentions. We argue here that these two sources of behavior are in fact not substitutable. Therefore, if α and β are measured in a context where various motives for social preferences are involved, then the parameters aggregate these motives quantitatively.

In formal terms, we argue that α_i and β_i estimates from the *inequity aversion list*

[19]The order of these tasks was randomized, the details of the randomization are further explained in section 1.4 *Experimental design*.



(a) Alpha (envy) (b) Beta (guilt)

Figure 1.6: Parameters' distributions: List 1 vs. Games

Note: This figure includes participants weakly adhering to Fehr and Schmidt's model. N=80. 40 participants experienced the lists first and 40 participants the games first (There are not significant differences between these two groups.) The numbers on top of the dark gray bars indicate the average for the correspondent range. Light gray bars do not have averages on top because based on the games we estimate bounds for the parameters, not point estimates.

are smaller than the estimates from games: $\hat{\alpha}_i$ and $\hat{\beta}_i$. Games' estimates correspond to $\hat{\alpha}_i = \alpha_i + \eta_i$ and $\hat{\beta}_i = \beta_i + \zeta_i$, where η_i and ζ_i capture the role of intentions.

Other Motives

By construction, the *Inequity Aversion List* prevents the confounding effect of motives as own-income and reciprocity that could bias the estimates in F&S. The former is alleviated by holding the decision maker's payoff constant and the latter by assigning the payoffs randomly and thus ruling out any *procedure-based* fairness concern. There are, however, other motives that may be driving the behavior we observe in the lab. This does not mean that fairness concerns and these other motives are necessarily mutually exclusive. Nonetheless, we can identify which motive is dominant over other potential motives. We explore those motives next.

The first motive we explore are maximin preferences. This preferences consist of a desire to maximize the minimal payoff in a group. Engelmann and Strobel (2004) find

that maximin preferences are one of the main drivers of behavior in an experiment where participants make choices over distributions allocating money among themselves and other participants in a group. An individual driven by maximin preferences would have a distinct pattern of response to our *Inequity Aversion List*: her utility would increase as individual j's payoff increases when inequality is advantageous ($x_i > x_j$), and then it would be fixed at a maximum point for all other allocations ($x_i \leq x_j$).

Note that maximin preferences fit our definition of weak adherence to F&S. However, we only observe this pattern in the responses given by 3 individuals, i.e. 3.7% of all weakly adherent individuals. Therefore, we conclude that maximin preferences are not driving the high rate of adherence to F&S that we observe for individuals in our sample. Figure C.8 in the appendix shows the results for individuals driven by maximin preferences.

The second motive we explore is social surplus maximizing behavior, also known as efficiency concerns, where an individual focuses on maximizing the combined payoff of all individuals. An individual motivated by efficiency concerns would display an upward trend pattern of response throughout the list. Note that in the *inequity aversion list* the decision maker's payoff is fixed at \$20, while the other person's payoff is changing from row to row. An increasing utility in the first part of the list, i.e. when $x_j \in [\$10, \$20]$ is consistent with both inequity aversion concerns and efficiency concerns. Fortunately, the second part of the list, when $x_j \in [\$21, \$30]$, can help us distinguish inequity aversion from efficiency concerns. On one hand, an individual whose behavior is driven by social surplus maximization has an increasing utility throughout the entire list. On the other hand, an individual driven by inequity aversion has an increasing utility only in the first part of the list. Therefore, social surplus maximizers don't fit either definition of adherence to F&S and thus efficiency concerns are not driving our results. Moreover, we only observe 3 individuals with dominant surplus maximizing behavior in our sample. Figure C.9 in the appendix depicts these subjects' results.

Evidence of the greater importance of inequity aversion in explaining subject's behavior in our sample contributes to a wider literature that discusses the relative importance of

inequity aversion, social surplus maximization and maximin preferences (See Andreoni & Miller, 2002; Charness & Rabin, 2002; Güth, Kliemt, & Ockenfels, 2003; Engelmann & Strobel, 2004; Fehr, Naef, & Schmidt, 2006; Daruvala, 2010; Galeotti, Montero, & Poulsen, 2018).

What are the motives displayed by non-adherents to F&S? We observe a general trend of decreasing utility on the other person's payment (See Figure C.7 on the appendix). This behavior is in line with the relative income hypothesis, which we explore in more detail in a separate paper (Zarghamee, Ifcher, Houser, & Diaz, 2020). Out of the 26 non-adherents to the F&S model in our sample, the general trend displayed by 22 participants is considered as envious or spiteful according to Kerschbamer's (2015) archetypes. One participant displays a pattern similar to inequity aversion, except that she maximizes her utility at an allocation of $20 for herself and $21 for the other, instead of equality. The remaining 3 participants are the aforementioned subjects who display social surplus maximizing behavior, their behavior is considered as either maximin or altruistic according to Kerschbamer's (2015) archetypes.

1.5.3 Demographic Characteristics

In the last part of the experiment we collect data about subject's demographic characteristics. Here we use that data to test whether certain characteristics are correlated with an individual's likelihood of displaying inequity aversion. Also, conditionally on displaying inequity aversion, we test whether parameters in F&S are systematically different depending on certain characteristics.

We use a probit model to test whether certain demographic characteristics have an effect on model adherence and we find no consistent statistically significant results for any of the variables. Table B.1 in the appendix, summarizes results for regressions estimating the likelihood of model adherence depending on gender, age, parent's education attainment, political views, majoring in Economics and GPA[20].

[20]We ask participants about their income level in two ways: we ask them to report their best estimates of their total expenditures during the school year, and the total income of their parents or guardians during the last year. Since we find a low correlation between these two variables, we decided to refrain from using income as a control variable in our analysis for lack of reliability.

Fehr et al. (2006) find that women and non-economists[21] choose egalitarian allocations more often than men and economists, respectively, at the expense of total social surplus. In contrast, they find that political attitudes and age don't affect social preferences for efficiency and equity.

Using Wilcoxon rank-sum tests, to account for the small sample size, we check whether the guilt and envy coefficients differ depending on the same characteristics tested by Fehr et al. (2006)(i.e. women, non-economists[22], individuals with different political views or different age groups) and two additional characteristics: student's GPA and parent's education. We find no statistically significant differences with respect to the guilt and envy coefficients for any demographic characteristic. Table B.2 in the appendix shows the results.

In the case of non-economists versus economists, our results are in line with Daruvala (2010), who also does not find differences between these two groups. This indicates that even though students of economics and business administration might have been taught about the importance of efficiency in their classes, they did not value it more than students in other fields.

To summarize, we don't find any effect of demographic characteristics on adherence to F&S nor on the parameters' size. Nonetheless, these results should be not taken as conclusive, since our tests may be under-powered due to a small sample size.

1.5.4 Robustness Checks

One potential concern regarding our methodology is that participants provide random answers to the *inequity aversion list* given that their expressed utility does not impact the likelihood of certain allocation being chosen for payment. However, evidence from the second part of the experiment suggests that participants in our experiment do not answer randomly. In the second part of the experiment, participants are asked to replicate their answers to List 1, part 1, and they are paid based on accuracy.

Figure 1.7 shows that more than 60% of all answers to List 1 given by participants in

[21] In this case economists are Economics and Business administration undergraduate students.
[22] defined as in Fehr et al., 2006.

the second part of the experiment are within 0.5 units away from the answers provided in the first part of the experiment for the corresponding allocation[23]. This level of consistency between the first and the second part gives us confidence that participants completed List 1 conscientiously in the first part of the experiment.

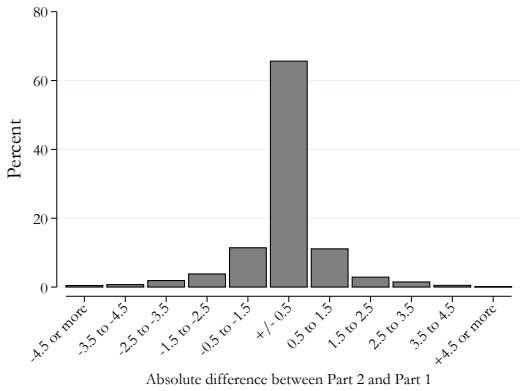


Figure 1.7: Absolute change in List 1 responses

Note: Histogram of the absolute change between the second part and the first part of the experiment, defined as the difference between the utility of an allocation in List 1 Part 2 and the utility of the same allocation in List 1 Part 1. This graph includes the 21 allocations in List 1 for all 106 participants, for a total of 2226 observations. Given that the utility is measured in a scale from 1 to 7 the maximum possible difference is 6.

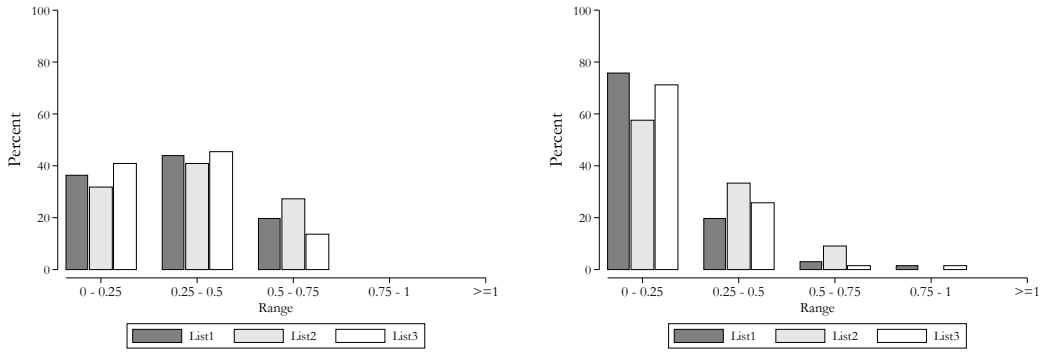
To further test whether participants provide reliable answers, we conduct a Monte Carlo simulation, generating random data and comparing it with the answers provided by our participants to List 1, part 1. The responses provided by our participants are statistically significantly different from random values. Please refer to Appendix D to see graphs summarizing the results and some specific details related to the simulations.

In addition to testing whether participants provide random responses to List 1, we conduct robustness tests to check whether modifications to the list change the parameters captured with our methodology. Lists 2 and 3 introduce modifications to List 1 in two different ways. List 2 changes the relative inequality of each allocation by subtracting $10

[23]We show an equivalent graph in terms of percentage change in figure C.10 in the appendix.

from every payoff in List 1. List 3 extends List 1 including allocations with larger inequality in both directions: advantageous and disadvantageous.

We obtain similar results at the aggregate level with Lists 2 and 3 than with List 1. Figures C.4 and C.5 in the appendix replicate figure 1.4 using List 2 and List 3 respectively. In each case, we find strong adherence to F&S. At the individual level, 84 participants weakly adhere to F&S using List 2, and 80 using List 3, compared to 80 using List 1. 65 participants strongly adhere to F&S using List 2, and 61 using List 3, compared to 69 using List 1.



(a) Alpha (envy)

KS p-value List1 vs. List2 = 0.717
KS p-value List1 vs. List3 = 0.435

KS p-value List2 vs. List3 = 0.154

(b) Beta (guilt)

KS p-value List1 vs. List2 = 0.103
KS p-value List1 vs. List3 = 0.852

KS p-value List2 vs. List3 = 0.318

Figure 1.8: Parameters' distributions: Lists 1, 2 and 3

Note: This figure includes participants displaying weak adherence to Fehr and Schmidt's model when completing all three lists. N=66.

Figure 1.8 depicts the distributions of α_i and β_i elicited with each list. The results of Kolmogorov-Smirnov tests for equality of distributions are presented below each graph. The hypothesis that the distributions are statistically equal is not rejected in any of the paired comparisons.

Although statistical equality of distributions is a good signal of the robustness of the *inequity aversion list*, we are interested in testing the consistency of our measure across

lists at the individual level. Therefore, we implement a more stringent test. Consider the following two equations, where m and n correspond to one of the three lists with $m \neq n$.

$$\alpha_{im} = \gamma_0 + \gamma_1 \alpha_{in} + \epsilon \tag{1.5}$$

$$\beta_{im} = \gamma_0 + \gamma_1 \beta_{in} + \epsilon \tag{1.6}$$

Equations (1.5) and (1.6) describe the relationship between parameters estimated using list m and parameters estimated using list n for the same group of individuals. Note that a scenario in which a parameter estimated using list n is a good predictor of the same parameter using list m implies that $\gamma_0 = 0$ and $\gamma_1 = 1$. Therefore, to test the consistency of our measure across lists we conduct a two step process: fist, we estimate the regressions described by equations (1.5) and (1.6); second, we conduct a post-estimation test to check whether $\gamma_0 = 0$ and $\gamma_1 = 1$, which corresponds to the following null hypotheses.

$$H_0 : \overline{\alpha}_{im} = \overline{\alpha}_{in} + \epsilon \tag{1.7}$$

$$H_0 : \overline{\beta}_{im} = \overline{\beta}_{in} + \epsilon \tag{1.8}$$

These null hypotheses express that on average the parameters elicited with list m are statistically not different from the parameters elicited with list n. We use a tobit model when estimating (1.5) and (1.6) to consider the left censoring of α_i, and the left and right censoring of β_i. Then, we test the null hypotheses by jointly testing whether $\gamma_0 = 0$ and $\gamma_1 = 1$ using an F-test.

Our results indicate that the null hypotheses are not rejected for $\gamma_1 = 1$ and for certain values of γ_0 close to 0. Table B.3 in the appendix summarizes the range of values for which the null hypotheses are not rejected. Although in some cases the range does not contain $\gamma_0 = 0$, the values are always close to 0. We conclude that variations to List 1, as those presented in lists 2 and 3, are consistent with List 1. Therefore the *inequity aversion list* is robust to changes to the list's parameters. This constitutes an improvement upon the

methodology proposed by Yang et al. (2016), which is sensitive to changes in the values of their lists, as pointed out by He and Wu (2016). Perhaps the robustness of our measure stems from using several data points to calculate F&S parameters with our methodology, as opposed to only one switching point with Yang et al.'s (2016) methodology.

1.6 Discussion

Heterogeneous social preferences have proven to be good predictors of behavior outside competitive markets. Several models have incorporated such preferences into the analysis of individual's behavior, improving predictions in diverse environments. These environments include: firms and organizations; markets with a small number of traders or with informational frictions, and environments where contracts are neither completely specified nor enforceable (Fehr & Schmidt, 2006). However, there is still room to refine the measure of social preferences and consequently improve our understanding of human behavior outside competitive markets. In this paper we contribute to the improvement of inequity aversion measures as proposed by Fehr and Schmidt's (1999) model.

Although we focus on the role of inequity aversion in the context of F&S, efficiency concerns could also explain some of the behavior observed. However, we find no support for social surplus maximization. Only 3% of participants in our sample display behavior consistent with efficiency concerns. This result is aligned with Fehr et al. (2006) who show that, in general[24], inequity aversion motives have a larger weight on individual's behavior than efficiency motives.

The case of maximin preferences is very especial because an individual driven by maximin preferences could be identified by our methodology as weakly adherent to F&S, although not strongly adherent. This would not be the case for subjects driven by efficiency concerns, as their preferences wouldn't be considered consistent with F&S. For participants

[24]Engelmann and Strobel (2004) find that efficiency concerns and maximin preferences rationalize Economics and Business undergraduates behavior better than inequity aversion. Fehr et al. (2006) include students from other disciplines and nonacademic employees. They conclude that the dominance of efficiency concerns over inequity aversion is restricted to Economics and Business students. For non-economists inequity aversion weighs more than efficiency concerns.

in our sample, not all the behavior characterized as weakly consistent with F&S is indistinguishable from maximin preferences. In fact, only 3% of subjects in our sample displayed behavior consistent with both inequity and maximin preferences. This finding runs opposite to Engelmann and Strobel (2004) who find that a combination of maximin preferences and efficiency concerns explain behavior better than inequity aversion, for individuals in their sample.

1.7 Conclusion

Estimations of Fehr and Schmidt's model parameters have shown an incongruity between behavior at the aggregate and the individual levels. We argue that the cause of the incongruity stems from the use of certain standard games, i.e. ultimatum and dictator, to estimate the model parameters. Although deviations from the self-interest model in these games can be mostly attributed to fairness concerns, the F&S model represents fairness concerns as self-centered inequity aversion, whereas the games may incorporate another type of fairness concern in the form of reciprocity and the associated role of intentions. In other words, procedure-based fairness concerns.

In this study we propose a new methodology to estimate the parameters of F&S that isolates inequity aversion from procedure-based fairness concerns. We call it the *inequity aversion list*. The model parameters are estimated at the individual level by eliciting individuals' satisfaction with different levels of inequality.

We find that participants in our experiment display the behavior described by F&S at both the aggregate and the individual levels, solving the incongruity found with previous methodologies. We argue that this is the result of our methodology's ability to isolate inequity aversion from alternative other-regarding concerns. Additionally, we find that individuals display a greater level of consistency with the model's assumptions than previous methodologies.

Robustness checks of our measure show that the *inequity aversion list* captures the envy and guilt parameters in the F&S model in a consistent way. Therefore, changes in the values

of the inequity aversion list do not affect the estimations.

For our next step, we want to test the predictive power of F&S model in environments where fairness concerns are clearly circumscribed to inequity aversion. In our opinion, this would be the best test of the predictive power of the model in terms of making an 'apples to apples' comparison. Later on, it would be interesting to find a way to augment the inequity aversion measure including additional behavioral motives. This would allow us to predict behavior in environments like the standard ultimatum and dictator games that incorporate various motives as drivers of other-regarding behavior.

Our contribution is both methodological and substantive. The *inequity aversion list* is an instrument readily available to apply both inside and outside of the laboratory. It could potentially be combined with other measures of social preferences to strengthen individual predictions or it could be used as a measure of fairness concerns, to correct model misspecification in contexts where social preferences matter.

Chapter 2: Model selection[1]

2.1 Introduction

Theories explaining pro-social behavior abound in Economics. These theories explain the rationale behind certain behaviors in which individuals not necessarily maximize their own profits, and instead consider others' profits when making decisions. Thus the various theories reflect fairness concepts and other ideals people value, besides their own monetary interest. These include, for example, inequity aversion (Loewenstein et al., 1989; Fehr & Schmidt, 1999; Bolton & Ockenfels, 2000), reciprocity and the role of others' intentions (Rabin, 1993; Charness & Rabin, 2002; Dufwenberg & Kirchsteiger, 2004; Falk & Fischbacher, 2006; Segal and Sobel, 2007, 2008), and efficiency concerns (Charness & Rabin, 2000).

These theories would have been of small value have they not been tested to check their validity, both individually and against each other. Comparing the predictive power of those models against each other is essential in order to understand the underlying motives of pro-social behavior at the aggregate and individual levels. The verdict from the theory comparison is that none of the models explains human behavior comprehensively. However, in spite of the predominant heterogeneity, some models have displayed greater predictive power than others. Engelmann and Strobel (2004) find that a combination of efficiency concerns, maximin preferences and selfishness can rationalize most of the data in simple one-shot distribution experiments. Daruvala (2010) also finds that efficiency concerns are important[2] and that among the inequity aversion models Bolton and Ockenfels (2000) performs better than Fehr and Schmidt (1999).

[1]This chapter is joint work with Dr. Daniel Houser.

[2]In their experiment, one third of the responders have strong concerns for the total surplus.

In this chapter, I present a new methodology for model comparison using the *inequity aversion list* from chapter 1. Although in principle the *inequity aversion list* was designed to test F&S's inequity aversion model, there are other behavioral models that can be displayed when answering the list. This is an advantage of the *inequity aversion list* as individuals' answers are unconstrained and thus they are able to reveal their true preferences.

2.2 Methodology

The possible behavioral models that individuals can display when answering the *inequity aversion list* are summarized in figure 2.1. The nine models studied here constitute the different configurations of preferences with equality as a pivotal point. There are three possible shapes the utility function takes for allocations to the left of equality (i.e. where individual i has a higher payoff than individual j): increasing, decreasing, and constant. Similarly, to the right of equality the utility function can take the same three shapes: increasing, decreasing, and constant. The combination of the possible shapes to the left and right of equality results in the 9 possibilities depicted in figure 2.1. Note that the only assumptions imposed for these models are continuity and local monotonicity, namely preferences are monotonic for disadvantageous inequality allocations and monotonic for advantageous inequality allocations, although not necessarily globally monotonic.

Let's discuss each behavioral model considered here, in the order it appears in figure 2.1. Each model in figure 2.1 is shown from the perspective of an individual i, in the space of another person's payoff x_j and i's utility U_i. For all the different levels of x_j, individual i's own payoff is fixed at $\overline{x_i}$. Dotted lines indicate equality of payoffs: $\overline{x_i} = x_j$

- The first three models depicted in figure 2.1 display different forms of inequity aversion, in the way proposed by Fehr and Schmidt (1999). Their model extends the self-regarding model incorporating a distaste for inequity in two directions: disadvantageous inequity (envy), and advantageous inequity (guilt). The first model in figure 2.1 is the most strict depiction of F&S, in which an individual suffers from both guilt

and envy. In the second one, an individual suffers only from envy and is indifferent to different levels of advantageous inequity. On the contrary, in the third model an individual suffers from guilt only and is indifferent to disadvantageous inequity. The third model also corresponds to the maximin model in which utility increases as the minimum payoff among the two subjects increases.

- The fourth model is the canonical self-regarding model according to which individuals maximize their own utility with regard to their own monetary payoff exclusively. This model is also classified here as a version of F&S, given that F&S's model is also consistent with self-regarding individuals if their model parameters take the values of $\alpha_i = 0$ and $\beta_i = 0$.

- The fifth model is the Relative Income Effect (RIE), where subjective well-being decreases with other's income. These type of preferences have been accounted for in other behavioral models as spiteful (Levine, 1998), or competitive preferences (Charness & Rabin, 2002). Similarly, capped RIE embodies status seeking preferences but only to the point of equality. After equality, individual i's utility is fixed at its lowest level for any level of disadvantageous inequality.

- Equity aversion (Fershtman, Gneezy, & List, 2012) is the opposite of F&S's strict version of inequity aversion. Another way to name this model is as a preference for inequity. It is expressed when individuals' reach their lowest utility at equality and their highest at the maximum levels of inequality. Houser and Xiao (2010) find evidence of inequality-seeking punishment using dictator games in the laboratory.

- The next model is efficiency concerns (Engelmann & Strobel, 2004), also called surplus maximization or social welfare preferences (Charness & Rabin, 2002), where individuals display a preference for Pareto improvements. In this case, as i's payoff is fixed she increases her utility as j's payoff increases. The theoretical basis for this model is founded on the utilitarian ideas by Bentham and J. S. Mill.

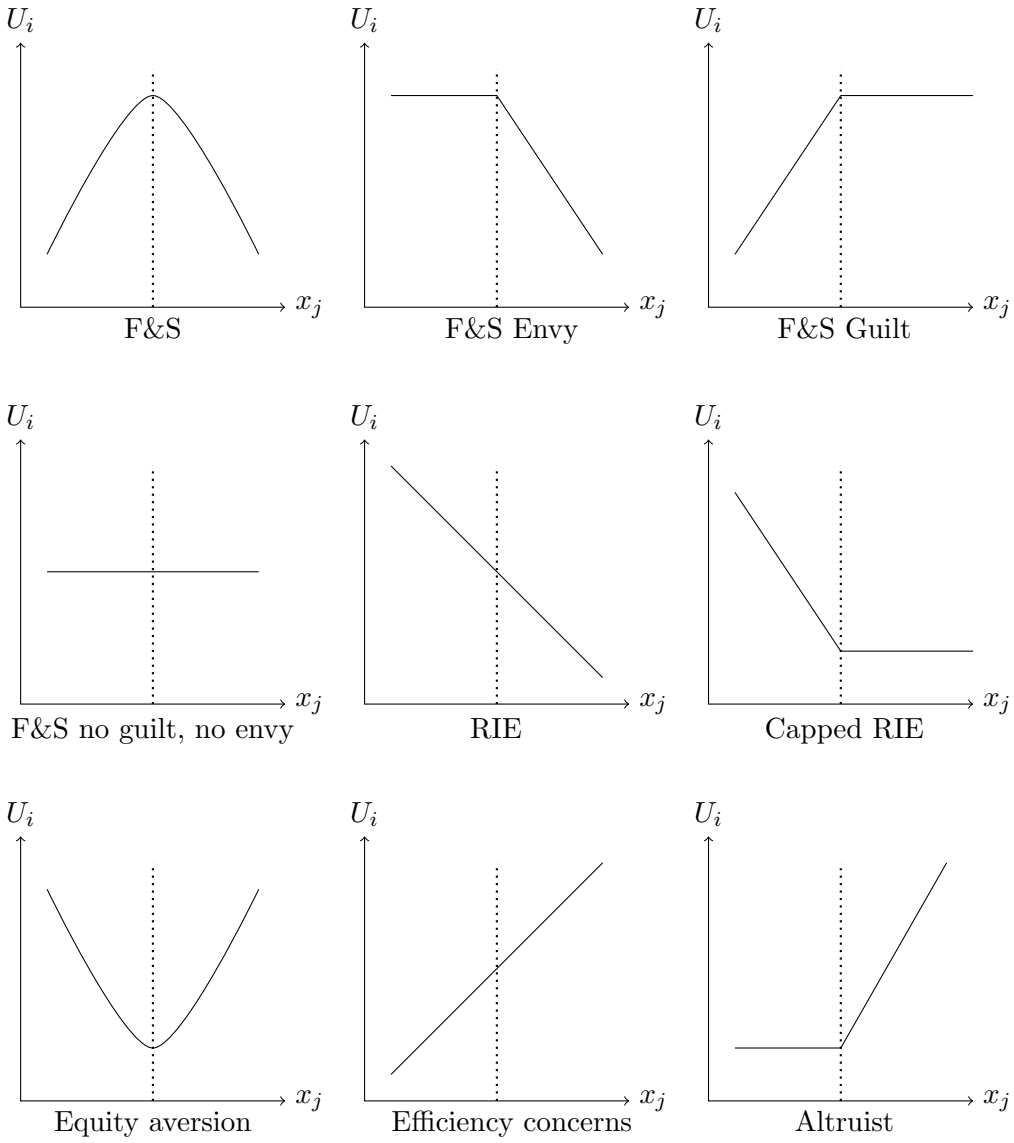


Figure 2.1: Possible behavioral models in the *inequity aversion list*

Note: Dotted lines indicate equality of payoffs: $x_i = x_j$

- The last model describes individuals who experience a low level of utility when their own monetary payoff is larger than that of other individual, while their utility increases with the other person's payoff once the other person gets at least the same amount.

Altruistic preferences are studied in the model proposed by Levine (1998).

2.2.1 Scoring Procedure

The next step, after defining the possible behavioral models displayed with the *inequity aversion list*, is to determine the model that fits each individual's behavior best. The approach we propose here is a scoring procedure that uses each allocation in the list as a single observation. The first observation is used as baseline and does not count towards the final score. For the second allocation, the utility an individual reports can be larger, smaller or the same as the first one. If the second allocation is larger than the first one, a unit is added toward the score of all the models increasing before equality (F&S, F&S Guilt, and efficiency concerns). If the second allocation is smaller than the first one, a unit is added toward the score of all the models decreasing before equality (RIE, capped RIE, and equity aversion). If the second allocation is the same as the first one, a unit is added toward the score of all the models that remain constant before equality (F&S Envy, F&S no guilt, no envy, and altruist). This process is repeated for each allocation in the list using as a reference the preceding allocation. Since the shape of most models changes after equality, the rules on how points are added adjust to the corresponding shape. For instance, after equality points are added towards F&S only if the utility reported for a given allocation is lower than for the preceding one.

After calculating the scores reached by a given individual for each model, the model of best fit for her preferences is that with most points. The *inequity aversion list* is composed of 21 allocations. Since the first allocation is used as a baseline, the maximum score a model can reach is 20 points.

2.3 Results

The results we analyse here correspond to the experiment described in the first chapter. In that experiment 106 participants completed the *inequity aversion list*. 13 individuals in our

sample have a perfect fit reaching a score of 20 points as the maximum among all models, and only 2 individuals have 9 points as the score for their best fit model. Tables B.4 and B.5, in the appendix, present the raw scoring results for each individual in our sample.

Among the possible behavioral models there were two that did not qualified as a model of best fit for any individual in our sample: F&S Guilt and Equity aversion. Table 2.1 summarizes the results for the rest of the models. The model that fits the majority of subjects (35.6%) is F&S Envy. The most important result is that on the aggregate, F&S is the model that best describes the behavior in our experiment: F&S envy, F&S no guilt, no envy, and F&S in its strict version account for the behavior of 84.7% of the subjects in our sample.

Previous studies have embarked on the task to discover the best behavioral models to rationalize behavior in economic environments. We compare our results with previous results with the caveat that it is unfeasible to make a perfect comparison. First, the set of actions available varies according to the environment studied, i.e. there are preferences we are able to observe with our list that we would not be able to observe in a dictator game. Second, as a consequence of the first reason, the set of behavioral models studied varies depending on the methodology, e.g. similarly to our study, Kerschbamer (2015) proposes a test that classifies individuals into nine preference types, however only seven of our behavioral models overlap with Kerschbamer's.

Andreoni and Miller (2002) find that 22.7% of subjects in a modified dictator game behave perfectly selfishly and 14.2% display Leontief preferences, the equivalent of what we denominate F&S guilt. These findings are in contrast with our findings, as we do not find evidence of F&S guilt. Similarly, only 6.8% of participants in our sample display efficiency concerns, while Charness and Rabin (2002) finds that efficiency concerns play a bigger role than inequity aversion for their sample. Similarly, Engelmann and Strobel (2004) find that efficiency concerns, maximin preferences, and self-regarding preferences rationalize most of their data.

Table 2.1: Model of best fit for each participant

Model	Frequency	Percentage
F&S Envy	42	35.6%
F&S no guilt, no envy	36	30.5%
F&S	22	18.6%
Efficiency concerns	8	6.8%
RIE	7	5.9%
Altruist	2	1.7%
Capped RIE	1	0.8%
Total	118	

Note: Although the number of subjects who participated in the experiment is 106, some subjects had a tie between models. The following is the list of ties: Three ties between Envy and RIE. A tie between F&S and Maxi-min. Five ties between F&S and Envy. A tie between capped RIE and Self-regarding. Two ties between Envy and Self-regarding

2.4 Conclusion

In this chapter we present a set of nine behavioral models that individuals can display when answering the *inequity aversion list* described in chapter 1. Then, we propose a scoring procedure to determine the model of best fit for each individual's preferences. Our results indicate that F&S outperforms other models in describing individual preferences over monetary allocations. However, the inequity aversion displayed by individuals in our sample concentrates on envy rather than guilt. In fact, the vast majority of subjects do not display guilt as an important part of their preferences.

Chapter 3: The relative income effect: an experiment[1]

3.1 Introduction

John Stuart Mill is credited with the claim that "men do not desire merely to be rich, but richer than other men."[2] This claim raises two fundamental questions. Does being richer than others make people happy? And if so, why? Empirical evidence regarding the first question is mixed. While subjective well-being (SWB) scholars have assembled substantial empirical evidence from large observational datasets of a negative relationship between others' income and one's own SWB (Blanchflower & Oswald, 2004; Clark, Frijters, & Shields, 2008; Helliwell, Layard, & Sachs, 2012; Luttmer, 2005), identification is often confounded. For example, some studies have found that cost-of-living explains the negative relationship (e.g., Ifcher, Zarghamee, & Graham, 2018), while others have estimated a positive relationship (Brodeur & Flèche, 2019; Clark, Kristensen, & Westergård-Nielsen, 2009; Deaton & Stone, 2013; D'Ambrosio & Frick, 2012; FitzRoy, Nolan, Steinhardt, & Ulph, 2014; Ifcher et al., 2018; Kingdon & Knight, 2007; Senik, 2004; Senik, 2008). A simple answer to the second question is that people care about relative consumption—that is, they are made happy by consuming more than others. Another plausible explanation that has not been explored in the SWB literature, though, is that people are made happy by comparing favorably to others in a more general sense, and being richer than others proxies for this ineffable relativity. In this case, being richer than others may only appear to increase happiness in the absence of other measures of relativity.

[1]This chapter is joint work with Dr. John Ifcher, Dr. Daniel Houser, and Dr. Homa Zarghamee. Financial support from Barnard College is gratefully acknowledged.

[2]Pigou (1920) attributes this quotation to Mill (Luttmer, 2005), but its authorship is contested by Mill scholars (see Rees (1956)).

In this paper, we attempt to address these two questions using an experiment. Specifically, we measure subjects' SWB before and after an exogenous relative-rewards shock in which subjects learn the number of experimental points they and another subject receive (2 or 10), and whether each experimental point is monetized (worth $1) or non-monetized (worth $0). This design enables us to compare the SWB-change of two subjects receiving the same experimental points who differ only in the points they learn another subject receives. Comparison of the impact of monetized and non-monetized points allows us to identify the impact of others' money on one's SWB and to determine whether non-monetized points generate the same observed relativity. To illustrate the nuance of this approach, consider a less nuanced version wherein a subject learns that she receives $2 and another subject receives $10. SWB may change because the subject learns that she is receiving $8 less than the other subject, but it may also be affected by learning that: (i) she is receiving $2; (ii) the other subject is receiving $10; (iii) there is a metric on which she is in worse relative standing than the other subject by 8 units; or (iv) she is receiving $4 less than she expected when agreeing to participate in the experiment. Our design attempts to rule out explanations (i) - (iv).

First, we find that it is not SWB-improving to learn that the other subject receives a smaller—in comparison to an equal—number of monetized points, suggesting that subjects are not made happier by being "richer" than other subjects ceteris paribus. Second, we find that while it is SWB-diminishing to learn that the other subject receives a larger—in comparison to an equal—number of monetized points, it is statistically indistinguishably SWB-diminishing when points are non-monetized. This suggests that others' money impacts own SWB only insofar as it proxies for others' points; that is, subjects are only made less happy about being "poorer" because it means they receive fewer points than others, not because they receive less money than others. In contrast, subjects do seem to be made happier by receiving money themselves, and not just for the points that the money represents: while it is SWB-improving to receive non-monetized points, it is significantly more SWB-improving to receive monetized points.

Our research is also motivated by the increasing use of SWB measures in public policy. In an influential critique of GDP as the singular measure of economic health and well-being, the Commission on the Measurement of Economic Performance and Social Progress recommended that policy makers consider SWB data (Stiglitz, Sen, Fitoussi, et al., 2009). Researchers have since been trying to understand the nature and limits of SWB indices as measures of well-being suitable for policy (e.g., Benjamin, Cooper, Heffetz, & Kimball, 2019). For example, Benjamin et al. (2012) conduct an experiment in which subjects read scenarios, predict their SWB in the scenarios, and make hypothetical choices between the scenarios; Benjamin, Heffetz, Kimball, and Rees-Jones (2014) compares medical students' real rankings of residency programs with their predicted SWB in those programs; and Benjamin, Heffetz, Kimball, and Szembrot (2014) estimates marginal utilities of various SWB measures and aspects of well-being. By testing whether and which SWB measures are responsive to experimental rewards, our experiment furthers inquiry into whether the experimental setting can offer for SWB policy researchers the combination of measurable outcomes and control that it offers choice-based researchers; as in all experiments, such lab-appropriate measures—along with any relativity that we identify in the experimental setting—should not be assumed to abide by the same dynamics observed in survey-based data.

3.2 Literature Review

The use of SWB data by researchers and policy-makers has proliferated in recent years: several national governments (e.g., Britain, France, and Bhutan) now collect SWB data from their citizens, and the World Bank now compiles SWB research and statistics in an annual World Happiness Report. SWB measures are numerical indices of respondents' sense of well-being and can range from items like, agreement with the statement "Did you feel worried for a lot of the day yesterday?" to agreement with "I lead a purposeful and meaningful life." Indeed, researchers have classified three broad categories of SWB measures: experiential-SWB measures assess feelings, experiences, or emotions over short

time-frames, like the present moment or in the last week; evaluative-SWB measures assess overall or domain-specific satisfaction or happiness with life; and eudemonic-SWB measures assess the fulfillment of underlying psychological needs, like respect, autonomy, and self-acceptance. The three measures are sometimes similarly responsive to potential correlates of SWB and sometimes not. For example, while a positive cross-sectional relationship between income and all three dimensions of SWB is well-established in the literature (e.g., Stevenson & Wolfers, 2008; Kahneman & Deaton, 2010; Clark & Senik, 2011), Kahneman and Deaton (2010) find that the relationship has a satiation point of $75,000 when using experiential but not evaluative measures; Stone and Mackie (2013) find that the relationship between SWB and having children is negative using experiential measures but positive using evaluative measures. Dolan, Layard, and Metcalfe (2011) discusses the policy-importance of separately measuring experiential, evaluative, and eudemonic SWB.

Often using evaluative measures, but also with experiential measures, SWB scholars have found evidence of a negative relationship between others' income and one's own SWB that is attributed to a "relative income effect" (RIE): income comparisons cause SWB to decrease with others' income, *ceteris paribus*.[3] The RIE has important implications, for example, regarding the potential benefits of economic growth. Some SWB researchers believe that the RIE helps explain the Easterlin Paradox: the empirical observation that over time in many countries, average national SWB does not increase with real per capita GDP (Easterlin, 1974; 2010; Easterlin, 2013). Luttmer (2005) reports a negative relationship between regional median income and SWB that is as big as the positive relationship between own income and SWB, implying that shared economic growth would not be associated with improved national SWB.

Identification of the RIE, though, is often confounded. First, observational studies do not generally account for selection (e.g., into neighborhoods or occupations). Indeed, the studies

[3]What we here call the RIE should not be mistaken for the Relative Income Hypothesis (RIH). While the RIE has SWB as its outcome variable, RIH, originally put forth by Duesenberry (1949), is the theory that consumption and savings *behavior* depend on one's income level relative to others'. It is technically ambiguous whether our motivating quote from Mill relates more to RIE or RIH. We speak further to the relationship between SWB and behavior in the discussion.

with the most credible exogenous relative-income shocks—the Moving to Opportunities for Fair Housing demonstration and the Dutch Postcode Lottery—do not find a negative effect of neighbors' income on one's SWB (Ludwig et al., 2013; Kuhn, Kooreman, Soetevent, & Kapteyn, 2011). Second, *ceteris paribus* is easily violated in observational studies. Ifcher et al. (2018) find that adding controls for cost-of-living renders the significant negative relationship between others' income and own SWB insignificant, suggesting that the original significant relationship could be explained by cost-of-living. Further, the sign of the others-income-own-SWB relationship is positive in some contexts. A positive relationship has been identified in immediate neighborhoods and has been attributed to local public goods (Ifcher et al., 2018; Brodeur & Flèche, 2019; Deaton & Stone, 2013[4]; Ludwig et al., 2013; Clark, Westergård-Nielsen, & Kristensen, 2009) or to altruism (Kingdon & Knight, 2007). A positive relationship has also been identified in periods of rapid economic growth and has been attributed to "the tunnel effect," whereby others' income serves as a signal of one's future income (Hirschman & Rothschild, 1973; Senik, 2004; 2008).

These violations of *ceteris paribus* illustrate that others' income can impact SWB through channels other than the RIE. They do not, however, prove or disprove the existence of the RIE itself, as the sign of the relationship between others' income and own SWB represents a net effect of an indeterminate set of channels. For example, Clark, Westergård-Nielsen, and Kristensen (2009) find that, controlling for own income, neighborhood median income and own SWB are positively correlated, which the authors attribute to local public goods. At the same time, controlling for own income and neighborhood median income, they find that one's income-rank within one's neighborhood is positively correlated with own SWB, which they attribute to the RIE.

These confounds render compelling the control afforded by an experimental analysis. To our knowledge, the only such experiment with real-money rewards is McBride (2010), who attempts to identify the RIE in a single-player, multiple-round, penny-matching game played against a computer. In each round, subjects are first informed of the computer's

[4]Deaton and Stone (2013) estimate a negative relationship using experiential measures and a positive relationship using evaluative measures.

randomized probability of choosing heads or tails (there are five possible types), and then they choose heads or tails. Next, subjects learn their own payment in that round and, depending on treatment, either: (i) no further information, (ii) the average payment of all other subjects, or (iii) the average payment of subjects by computer-probability-type. Subjects then report their satisfaction with the results of the round. Satisfaction with a round's results significantly decreases with the average payment of other subjects in (ii) and only with the average payment of subjects with the same probability-type in (iii). While these results appear consistent with the RIE, identification is confounded because subjects have agency: their satisfaction (dissatisfaction) may not be due to income per se but to feelings of relative success or self-congratulations (failure or self-criticism).[5]

3.3 Experimental Design

3.3.1 Discovery and pilot studies

Like the main online experiment, the goal of the discovery study was to identify the SWB-impact of an exogenous relative-rewards shock. A more general ambition was to determine whether the control afforded by laboratory experimentation could be leveraged to study SWB (as measured in large observational datasets and increasingly in national accounts) and its determinants, especially income. The discovery study was necessary to determine SWB measures' sensitivity to money-rewards of the size paid in experiments. This required testing the many different measures of SWB available in the literature, and varying payment-sizes. As such, the discovery study, while similar in methodology to the online experiment, includes much longer SWB surveys and more rewards-shocks.

[5]It warrants mention that other experiments have found evidence supportive of the RIE using choices over hypothetical scenarios (Johansson-Stenman, Carlsson, & Daruvala, 2002; Solnick & Hemenway, 1998; Zeckhauser, 1991). Also supportive of the RIE, Smith, Diener, and Wedell (1989) document higher SWB with a given hypothetical wage when it is from a distribution with a lower mean. Further, there are three economic experiments in which relative real-money payments vary and SWB is measured. The correlation between others' payments and subjects' SWB is consistent with the RIE in Bosman and Van Winden (2002) and Konow and Earley (2008) and inconsistent with the RIE in Charness and Grosskopf (2001). That said, these experiments are not designed to test the RIE so attempts to attribute the correlations to the RIE are beset by endogeneity and other identification issues.

While the discovery study enabled us to refine the SWB survey and revelation mechanism, the pilot study enabled us to test the revised methodology and help estimate the number of subjects that were needed for the online experiment.[6]

3.3.2 Online experiment

To study the RIE, we conducted an online experiment in December 2017. Prospective subjects were recruited using the online crowdsourcing marketplace Amazon Mechanical Turk (mTurk). The set of potential subjects consisted of mTurk "workers"—people seeking to make money by performing tasks on the platform—whose accounts were registered in the United States and had at least 95% of their past tasks approved by employers. The description of the task on mTurk informed potential workers (subjects) that participation would take less than 20 minutes, that they would be paid for their participation (minimum, average, and maximum payments of \$3.00, \$7.00, and \$11.00), and that the payment-amount would be determined randomly. Each recruited subject received a link to Qualtrics, where she received an exogenous relative-rewards shock (revelation) and completed pre- and post-revelation SWB surveys, allowing for measurement of the SWB-effect of the revelation. Subjects entered a survey-completion code generated by Qualtrics into mTurk to receive their payments. 996 subjects completed the experiment in 10 minutes on average and were paid a minimum, average, and maximum of \$3.00, \$7.27, and \$11.00.[7]

Pre- and post-revelation SWB surveys

The pre- and post-revelation SWB surveys each included the Mood Short Form (MSF), which enables a quick measurement of subjects' experiential SWB (Peterson & Matthew, 1983). The MSF includes the following four items, with a five-point Likert response-scale for each item ("Strongly disagree" =1, "Disagree" =2, "Neither agree nor disagree" =3,

[6]Columns (1) and (2) of table B.12 present the results of pooling the pilot study and online experiments, using the methods presented in Section 3.4.2 below; the results are materially unchanged.

[7]Of 1,180 individuals who clicked the Qualtrics link, 128 did not complete the survey, and another 56 completed the survey but either entered an incorrect survey-completion code or none at all. These individuals could not be paid and are not included as subjects in our analyses.

"Agree" =4, and "Strongly agree" =5):

- "Currently I am in a good mood."
- "As I answer these questions, I feel very cheerful."
- "For some reason I am not very comfortable."
- "At this moment I feel 'edgy' or irritable."

MSF scores are calculated by summing the four responses with the response-scale reverse-coded for the last two items. Possible scores range from 4 (worst possible mood) to 20 (best possible mood).

In an attempt to reduce the propensity of subjects to anchor their post- to pre-revelation MSF responses, items from the Basic Psychological Need Satisfaction Scale (BPNSS) were included in both surveys. The BPNSS is a 21-item instrument that measures needs for competence, autonomy, and relatedness (Deci & Ryan, 2000; Gagné, 2003). Responders evaluate statements (e.g., "Often, I do not feel very confident," and "People in my life care about me") on a seven-point scale in which "Not at all true" =1, "Somewhat true" =4, and "Very true" =7. The pre- and post-revelation SWB surveys included 11 and 10 BPNSS items, respectively.

To further reduce the propensity to anchor, the order of all items in the pre-revelation SWB survey was randomized by subject. The post-revelation items were randomized in two strata: the first contained the four MSF items randomized by subject, and the second contained the 10 BPNSS items also randomized by subject. The stratified randomization ensured that the post-revelation MSF items were presented immediately after the revelation, thus reducing concerns that the impact of the revelation may have worn off.

It is important to note why we use the MSF in the pilot and online experiments. For the discovery study, we compiled a comprehensive list of SWB items and scales that had been previously used in the SWB literature, including eudemonic, evaluative, and experiential measures; these came from oft-cited SWB datasets (e.g., the World Values Survey), and

contained popular SWB measures (e.g., Cantril's Ladder, a single-item measure) and scales (e.g., Positive Affect Negative Affect Scale, a multi-item measure). In the discovery sample, subjects completed a total of 92 SWB items (8 single-item and 7 multi-item measures), and, of these, the MSF was found to be the most responsive to revelation and the BPNSS among the least responsive.

Revelation Mechanism

The revelation mechanism was designed to provide an unconfounded relative-rewards shock. Prior to the pre-revelation SWB survey, in an attempt to standardize expectations, the instructions fully informed subjects about the revelation mechanism. Each subject was informed that:

- She would be paid a $1 reward for completing the study.[8]
- She would be randomly assigned to a two-person group.
- The other subject in the group (hereafter Participant X) could be any other subject in the study.
- She would never learn Participant X's identity and vice versa.
- She and Participant X would be allotted 2 or 10 experimental points each, creating four possible allocations of points:
  - She receives 2 points, and Participant X receives 2 points.
  - She receives 2 points, and Participant X receives 10 points.
  - She receives 10 points, and Participant X receives 2 points.
  - She receives 10 points, and Participant X receives 10 points.

[8] In mTurk, a flat payment that all subjects receive is called a "reward;" and a payment that can vary by subject is called a "bonus." It should be noted that our payments are sizeable for mTurk, where workers' median hourly rewards have been estimated to be $2, and only 4% of workers earn more than $7.25 per hour (Hara et al., 2018).

Below, we refer to these allocations as low-low (LL), low-high (LH), high-low (HL), and high-high (HH), respectively.

Subjects were then randomly informed of being in one of two treatments. Each subject in the "points-money (pts$) treatment" was informed that each experimental point was worth $1. Each subject in the "points (pts) treatment" was informed that she and Participant X would receive a $6 payment regardless of the number of experimental points. Subjects were presented with neither the allotment nor the treatment terminology. Also, subjects in the pts$-treatment did not know about the pts-treatment, and vice versa.

After completing the pre-revelation SWB survey, subjects were reminded, as appropriate, that either "the number of points you and Participant X receive will not affect your bonus payments," or "you and Participant X will receive a bonus payment of a dollar for each point you receive." Subjects were also reminded that they were randomly assigned to one of the four possible allocations of points. To reveal the allocation of points, subjects had to press an "OK" button. They were then informed of the number of points they and Participant X received.

Screening questions, questionnaire, and end of experiment

Though mTurk subjects have been shown to be more attentive than convenience samples from universities (Hauser & Schwarz, 2016), because our survey items were not individually incentivized, the experiment included four screening questions to allay any concerns about workers being inattentive or having programmed responses:

- After reading the instructions and before answering the pre-revelation SWB survey, subjects were asked what year it was; there were five possible responses: "1990," "2017," "2000," "2018," and "2019."
- After completing the pre-revelation SWB survey, subjects were presented with the following statement: "Currently, the year is 2025." The response scale was the same five-point Likert scale used with the MSF items.[9]

[9]To minimize the impact of these screening questions on subjects' MSF responses, we chose to use

- In the post-revelation SWB survey (after completing the MSF items and before completing the BPNSS items), subjects completed the following two items:
  - "Please indicate the value of the bonus you will receive (in addition to the $1 reward you will receive for completing this study)." Possible responses were "$2," "$6," and "$10."
  - "Please indicate the value of the bonus Participant X will receive (in addition to the $1 reward Participant X will receive for completing this study)." Possible responses were "$2," "$6," and "$10."

After the post-revelation SWB survey, subjects completed a 15-item questionnaire (including demographic characteristics and political views), entered their survey-completion codes, and were paid.

3.3.3 Own- and relative-rewards shocks

Our experiment yields 8 distinct cells from a 4 (allocations of 2 or 10 pts to the subject and 2 or 10 pts to Participant X) X 2 (pts$- versus pts-treatment) design: $LL_{pts\$}$, $LH_{pts\$}$, $HL_{pts\$}$, $HH_{pts\$}$, LL_{pts}, LH_{pts}, HL_{pts}, and HH_{pts} (see table B.6 in Appendix B). This design enables identification of both relative-rewards and own-rewards shocks by comparing cells that hold all other factors constant.

For example, let's compare the HL and HH cells by treatment. Each subject in the $HL_{pts\$}$ cell was told that she would receive 10 pts worth $1 each and that Participant X would receive 2 pts worth $1 each, whereas each subject in the $HH_{pts\$}$ cell was told that she and Participant X would each receive 10 pts worth $1 each. In these two cells, each subject's realized pts- and $-rewards are constant (10 pts and $10). Further, as the instructions disclosed all possible revelations, each subject's expected pts- and $-rewards should be 6 pts and $6 for both herself and Participant X. Thus, the $HL_{pts\$}$ and $HH_{pts\$}$ cells differ

questions about what year it was, as the questions' content is affectively neutral. We asked the two questions in two different ways to encourage attention, and the second question was designed as a statement with a Likert scale to mimic MSF items.

only in Participant X's realized pts- and \$-rewards (2 pts and \$2 versus 10 pts and \$10). By comparing the MSF-change (post- minus pre-revelation) of $HL_{pts\$}$ and $HH_{pts\$}$ subjects, we can identify *ceteris paribus* the differential impact on a subject's mood of Participant X receiving 2 pts and \$2 rather than 10 pts and \$10: $\Delta MSF_{HL,pts\$}$ - $\Delta MSF_{HH,pts\$}$. We refer to this as the impact of an advantageous relative-pts-\$ shock. In relative-rewards-shock comparisons, "advantageous" indicates that the subject receives greater rewards than Participant X, rather than the same rewards as Participant X.

Let's now consider the analogous cells of the pts-treatment. Because pts are not monetized, and the realized \$-rewards are constant in the HL_{pts} and HH_{pts} cells, $\Delta MSF_{HL,pts}$ - $\Delta MSF_{HH,pts}$ identifies *ceteris paribus* the differential impact on a subject's mood of Participant X receiving 2 pts rather than 10 pts. We refer to this as the impact of an advantageous relative-pts shock. Therefore, we can capture the impact of Participant X receiving \$2 rather than \$10 by subtracting the impact of an advantageous relative-pts shock from the impact of an advantageous relative-pts-\$ shock: ($\Delta MSF_{HL,\$}$ - $\Delta MSF_{HH,\$}$) = ($\Delta MSF_{HL,pts\$}$ - $\Delta MSF_{HH,pts\$}$) - ($\Delta MSF_{HL,pts}$ - $\Delta MSF_{HH,pts}$). We refer to this difference as the impact of an advantageous relative-\$ shock.

In relative-rewards-shock comparisons, "disadvantageous" indicates that the subject receives lesser rewards than Participant X, rather than the same rewards as Participant X. By comparing LH and LL cells, we can identify the impact of disadvantageous relative-pts-\$ ($\Delta MSF_{LH,pts\$}$ - $\Delta MSF_{LL,pts\$}$), relative-pts ($\Delta MSF_{LH,pts}$ - $\Delta MSF_{LL,pts}$), and relative-\$ (($\Delta MSF_{LH,\$}$ - $\Delta MSF_{LL,\$}$) = ($\Delta MSF_{LH,pts\$}$ - $\Delta MSF_{LL,pts\$}$) - ($\Delta MSF_{LH,pts}$ - $\Delta MSF_{LL,pts}$)) shocks.

Similarly, we can identify the impact of an advantageous own-pts-\$ shock by comparing $HH_{pts\$}$ to $LL_{pts\$}$ cells: $\Delta MSF_{HH,pts\$}$ - $\Delta MSF_{LL,pts\$}$. Subjects in the $HH_{pts\$}$ ($LL_{pts\$}$) cells are told that they and Participant X will each receive 10 pts (2 pts) worth \$1 each. In both cells, subjects and Participant X receive the same rewards, thus the RIE should not explain the impact of the advantageous own-pts-\$ shock. Therefore, in own-rewards-shock comparisons,

"advantageous" indicates that the subject receives greater rewards than expected, rather than lesser rewards than expected (recall that expected rewards are 6 pts and \$6 for all subjects).

Comparing the analogous cells of the pts-treatment, we can identify the impact of an advantageous own-pts shock by comparing HH_{pts} and LL_{pts} cells: $\Delta MSF_{HH,pts}$ - $\Delta MSF_{LL,pts}$. Because pts are not monetized and the realized \$-rewards are constant, subjects in these cells differ only in the pts-rewards that they and Participant X receive (10 pts each versus 2 pts each). Therefore, we can capture the impact of the subject receiving \$10 rather than \$2 by subtracting the impact of an advantageous own-pts shock from the impact of an advantageous own-pts-\$ shock: ($\Delta MSF_{HH,\$}$ - $\Delta MSF_{LL,\$}$) = ($\Delta MSF_{HH,pts\$}$ - $\Delta MSF_{LL,pts\$}$) - ($\Delta MSF_{HH,pts}$ - $\Delta MSF_{LL,pts}$). We refer to this difference as the impact of an advantageous own-\$ shock. In Panel A of Table B.7, we provide a complete list of the relative- and own-rewards shocks and, in Panel B, the formulae we use to identify their impacts.[10]

3.3.4 Experimental predictions

Relative-rewards-shocks predictions

Making no further assumptions, the RIE yields two unambiguous predictions about the impact of relative-rewards shocks:

- Positive impact of the advantageous relative-pts-\$ shock:
  - $\Delta MSF_{HL,pts\$}$ - $\Delta MSF_{HH,pts\$} > 0$
- Negative impact of the disadvantageous relative-pts-\$ shock:
  - $\Delta MSF_{LH,pts\$}$ - $\Delta MSF_{LL,pts\$} < 0$

These predictions can be inferred from the definition of the RIE, as both comparisons hold constant the subject's monetized pts and vary Participant X's, with smaller (larger)—in

[10]The impact of disadvantageous own-rewards shocks are not listed, as they can be calculated by multiplying the impact of the corresponding advantageous own-rewards shocks by negative one.

comparison to an equal—number of monetized pts for Participant X associated with higher (lower) values of ΔMSF.

Predicting the impact of the other relative-rewards shocks requires assumptions about MSF-scores' responsiveness to non-monetized pts. We will consider two alternative assumptions: (a) strict RIE, whereby the RIE applies to monetized but not non-monetized pts, versus (b) generalized relative effect (RE), whereby the RIE applies to monetized pts and an analogous relative effect—such that SWB is decreasing in others' pts, holding own pts constant—applies with equal magnitude to non-monetized pts. These assumptions represent two extreme and non-exhaustive cases—an intermediate case could be that non-monetized pts impact mood but not as much as do monetized pts.

Under the strict RIE, the impacts of both advantageous and disadvantageous relative-pts shocks are predicted to be zero, and this yields the prediction that \$-shocks will have the same impact as their corresponding pts-\$-shocks. Specifically:

- Zero impact of the advantageous relative-pts shock:
  - $\Delta\text{MSF}_{\text{HL,pts}}$ - $\Delta\text{MSF}_{\text{HH,pts}} = 0$
- Zero impact of the disadvantageous relative-pts shock:
  - $\Delta\text{MSF}_{\text{LH,pts}}$ - $\Delta\text{MSF}_{\text{LL,pts}} = 0$
- Equal impacts of the advantageous relative-\$ and advantageous relative-pts-\$ shocks:
  - $\Delta\text{MSF}_{\text{HL,\$}}$ - $\Delta\text{MSF}_{\text{HH,\$}} = \Delta\text{MSF}_{\text{HL,pts\$}}$ - $\Delta\text{MSF}_{\text{HH,pts\$}}$
- Equal impacts of the disadvantageous relative-\$ and disadvantageous relative-pts-\$ shocks:
  - $\Delta\text{MSF}_{\text{LH,\$}}$ - $\Delta\text{MSF}_{\text{LL,\$}} = \Delta\text{MSF}_{\text{LH,pts\$}}$ - $\Delta\text{MSF}_{\text{LL,pts\$}}$

Under generalized RE, the impacts of the advantageous and disadvantageous relative-pts shocks are predicted to equal the impacts of the corresponding pts-\$ shocks, and this yields the prediction that the relative-\$ shocks will have no impact.

- Equal impacts of the advantageous relative-pts and advantageous relative-pts-$ shocks:
  - $\Delta\text{MSF}_{\text{HL,pts}}$ - $\Delta\text{MSF}_{\text{HH,pts}} = \Delta\text{MSF}_{\text{HL,pts\$}}$ - $\Delta\text{MSF}_{\text{HH,pts\$}}$
- Equal impacts of the disadvantageous relative-pts and disadvantageous relative-pts-$ shocks:
  - $\Delta\text{MSF}_{\text{LH,pts}}$ - $\Delta\text{MSF}_{\text{LL,pts}} = \Delta\text{MSF}_{\text{LH,pts\$}}$ - $\Delta\text{MSF}_{\text{LL,pts\$}}$
- Zero impact of the advantageous relative-$ shock:
  - $\Delta\text{MSF}_{\text{HL,\$}}$ - $\Delta\text{MSF}_{\text{HH,\$}} = 0$
- Zero impact of the disadvantageous relative-$ shock:
  - $\Delta\text{MSF}_{\text{LH,\$}}$ - $\Delta\text{MSF}_{\text{LL,\$}} = 0$

Own-rewards-shock predictions

Assuming that subjects are made happier by receiving more rather than fewer monetized pts, and/or that, relative to expectations, subjects are made happier by gains of monetized pts in comparison to losses, the following prediction obtains:

- Positive impact of the advantageous own-pts-$ shock:
  - $\Delta\text{MSF}_{\text{HH,pts\$}}$ - $\Delta\text{MSF}_{\text{LL,pts\$}} > 0$

Additional predictions about own-rewards shocks require additional assumptions. If we assume that, analogous to the strict RIE, MSF-scores do not respond to own non-monetized pts, then an advantageous own-pts shock is predicted to have no impact, and an advantageous own-$ shock will have the same impact as an advantageous own-pts-$ shock:

- Zero impact of the advantageous own-pts shock:
  - $\Delta\text{MSF}_{\text{HH,pts}}$ - $\Delta\text{MSF}_{\text{LL,pts}} = 0$
- Equal impacts of the advantageous own-$ and advantageous own-pts-$ shocks:

- $\Delta\text{MSF}_{\text{HH,\$}}$ - $\Delta\text{MSF}_{\text{LL,\$}} = \Delta\text{MSF}_{\text{HH,pts\$}}$ - $\Delta\text{MSF}_{\text{LL,pts\$}}$

If we assume that, analogous to the generalized RE, MSF-scores respond equally to own non-monetized and monetized pts, then an advantageous own-pts shock is predicted to have the same impact as an advantageous own-pts-$ shock, and an advantageous own-$ shock will have no impact:

- Equal impacts of the advantageous own-pts and advantageous own-pts-$ shocks:
  - $\Delta\text{MSF}_{\text{HH,pts}}$ - $\Delta\text{MSF}_{\text{LL,pts}} = \Delta\text{MSF}_{\text{HH,pts\$}}$ - $\Delta\text{MSF}_{\text{LL,pts\$}}$
- Zero impact of the advantageous own-$ shock:
  - $\Delta\text{MSF}_{\text{HH,\$}}$ - $\Delta\text{MSF}_{\text{LL,\$}} = 0$

3.4 Results

Of the 996 subjects who completed the experiment, 471 were in the pts-treatment and 525 in the pts$-treatment. 132 subjects (69 in the pts-treatment and 63 in the pts$-treatment) completed at least one of the screening items incorrectly and were dropped from the analysis.[11] Table B.8 presents the demographic characteristics of the sample, comparable to other mTurk samples in the literature (see Kuziemko, Norton, Saez, and Stantcheva (2015) and Snowberg and Yariv (2018)). Table B.9 presents the pre- and post-revelation MSF-scores and ΔMSF for each cell.

[11] 2 subjects did not indicate that the year was "2017;" 9 subjects did not "Disagree" or "Strongly Disagree" that the current year was 2025; 53 (45) subjects indicated the incorrect payment for themselves (Participant X) in the pts$-treatment; and 56 (64) subjects indicated the incorrect payment for themselves (Participant X) in the pts-treatment. The main results, using the methods presented in Section 3.4.2 below, hold if we do not drop these subjects (see columns (3) and (4) of Table B.12). It can be argued that keeping these subjects in the analysis may be preferred, as it may be more reflective of a general population in which some individuals do not pay close attention to their own and/or others' income; thank you to Alan Kirman for pointing this out.

3.4.1 Non-parametric tests of own- and relative-rewards shocks

In Table B.10, we present the estimated impacts on ΔMSF of all own- and relative-rewards shocks.[12] Because we use multiple outcomes, we use Bonferroni adjustments to control for the familywise error rate; List, Shaikh, and Xu (2016) consider this the most conservative approach. The Bonferroni adjustments are applied to alphas, not p-values; so the p-values reported in Table B.10 are unadjusted. Bonferroni-adjusted alphas are equal to the standard Type 1 error rate (e.g., 0.05) divided by the number of tests in a family; starring in the table corresponds to whether the raw p-value is less than the Bonferroni-adjusted alpha. We define three families, each of which has three tests. The families are: advantageous own-rewards (first three rows of Table B.10), advantageous relative-rewards (next three rows of Table B.10), and disadvantageous relative-rewards (last three rows of Table B.10).

Considering the own-rewards shocks, we observe that the impact of advantageous own-pts-$, own-pts, and own-$ shocks are positive, statistically significant, and economically meaningful. For example, the impact of the advantageous own-$ shock is 2.17 MSF-points. In other words, controlling for the impact of own pts, subjects who receive $10 rather than $2 *ceteris paribus* experience a mood-improvement of roughly 14% of the pre-revelation sample-mean MSF-score (= 2.17/15.25). The impact of the advantageous own-pts shock being positive and significant indicates that subjects' mood is impacted by receiving non-monetized pts. The impact of the advantageous own-$ shock being positive and significant indicates that the impact on subjects' mood of monetized pts exceeds the impact of receiving equivalent non-monetized pts.

Turning to the advantageous relative-rewards shocks, there is support for neither the strict RIE, nor the generalized RE. Table B.10 shows that the impacts of the advantageous relative-pts-$ and relative-pts shocks are both negative, with the latter statistically significantly so. While the impact of the advantageous relative-$ shock is positive—consistent with the strict RIE insofar as it is less mood-diminishing to be "richer" than other subjects

[12] Columns (5) and (6) of Table B.12 show, using the methods from Section 3.4.2 below, that the relative- and own-rewards shocks did not impact BPNSS scores.

in both pts and $ than in pts alone—it is not statistically significant. Taken together, there is no evidence that it is mood-improving to be "richer" than other subjects. If anything, the results of the advantageous relative-pts shock are suggestive of it being mood-diminishing to be "richer" than other subjects, though we can only speculate as to why this is only the case for relative-pts and not relative-pts-$ shocks. It could be, for example, that concerns for social surplus maximization of monetized rewards lead to relative mood-improvement for those in $HH_{pts\$}$ (in which total social surplus is $20) relative to those in $HL_{pts\$}$ (in which total social surplus is $12).[13]

The estimated impacts of the disadvantageous relative-rewards shocks, though, favor the generalized RE. The impacts of the relative-pts-$ and relative-pts shocks are negative and significant. Further, they are statistically indistinguishable from each other, yielding an insignificant impact of the disadvantageous relative-$ shock. This suggests that being "poorer" than other subjects in monetized points is only as mood-diminishing as is being "poorer" in non-monetized points.

Comparing the impacts of the own-rewards and relative-rewards shocks reveals two interesting asymmetries. First, one's mood seems to depend on her own money—but not others' money—more than the points the money represents. Second, one's mood diminishes from being behind others but does not improve from being ahead.

3.4.2 Regression and subgroup analyses

To analyze the MSF-effects of relative-rewards shocks while controlling for demographic characteristics, we estimate the following equation:

[13]It should be noted that the results for disadvantageous relative-rewards shocks presented below are not consistent with a analogous concern for social surplus maximization, suggesting either that such concerns are not driving the advantageous relative-reward results or that social surplus concerns are less effective at tempering mood in the $LL_{pts\$}$ and $LH_{pts\$}$ conditions (in which subjects learn their payments are lower than expected) than in $HH_{pts\$}$ and $HL_{pts\$}$ (in which subjects learn their payments are higher than expected).

$$\Delta\text{MSF} = \beta_0 + \beta_1 I_{2\text{pts\$}} + \beta_2 I_{10\text{pts\$}} + \beta_3 I_{10\text{pts}} + \beta_4 I_{X10\text{pts}} + \beta_5 I_{2\text{pts\$},X10\text{pts\$}} + \beta_6 I_{10\text{pts\$},X10\text{pts\$}} + \beta_7 I_{10\text{pts},X10\text{pts}} + \delta Y + \epsilon \tag{3.1}$$

The indicator variable $\text{I}_{2\text{pts\$}}$ ($\text{I}_{10\text{pts\$}}$) equals one if the subject receives 2 pts (10 pts) worth \$1 each. The indicator variable $\text{I}_{10\text{pts}}$ (I_{X10pts}) equals one if the subject (Participant X) receives 10 pts. The indicator variable $\text{I}_{2\text{pts\$,X10pts}}$ ($\text{I}_{10\text{pts\$,X10pts\$}}$) equals one if the subject receives 2 pts (10 pts) worth \$1 each and Participant X receives 10 pts worth \$1 each. The indicator variable $\text{I}_{10\text{pts,X10pts}}$ equals one if the subject receives 10 pts and Participant X receives 10 pts. The vector Y contains demographic characteristics.[14] Robust standard errors are calculated.

The constant term β_0 captures the MSF-change experienced by subjects in the LL_{pts} cell. Regression coefficients can be used to recover the impacts of relative- and own-rewards shocks; the formulae are presented in Panel C of Table B.7 and the corresponding estimates are reported in Table B.11. Column (1) of Table B.11 excludes controls for demographic characteristics and replicates the corresponding results in Table B.10. Column (2) includes controls for demographic characteristics, yielding results similar to column (1).

Turning to subgroup analyses, we investigate all differences that are suggested by the SWB and social preferences literature for which we have demographic data in the questionnaire. Alesina, Di Tella, and MacCulloch (2004) find that conservatives' SWB is unaffected by—while progressives' SWB decreases with—income inequality in the US; this may be

[14] Demographic characteristics—all of which are treated as categorical variables in the regression analysis—include gender (female, male, other, prefer not to answer), age (bottom-quartile (< 29 years) and top-quartile (> 42 years)), race (American Indian/Alaska Native, Asian, Black/African American, Hispanic, Native Hawaiian/other Pacific Islander, White, other, prefer not to answer), religion (Atheist/Agnostic, Buddhist, Christian, Hindu, Jewish, Muslim, other, prefer not to answer), political leaning (conservative, liberal/progressive, moderate, prefer not to answer), household income (\$0-\$25K, \$25K-\$50K, \$50K-\$75K, \$75K-\$100K, \$100K-\$125K, \$125K-\$150K, \$150K+, prefer not to answer), education (some high school, completed high school, some college, completed college, some grad/professional school, completed grad/professional school, prefer not to answer), and employment status (employed, full-time; employed, part-time; not employed, looking for work; not employed, not looking for work; retired; student; prefer not to answer).

due to conservatives being more likely to attribute success to hard work and talent, and progressives to luck (Frank, 2016). Kamas and Preston (2015) identify gender-differences in social preferences, with women more likely to be inequity averse and men more likely to be social-surplus maximizers. Ifcher et al. (2018) and Alesina et al. (2004) find that the negative income-inequality-SWB relationship is significantly stronger for high- than low-income Americans, with no significant relationship for low-income subgroups in some specifications. Kuziemko, Buell, Reich, and Norton (2014) find that "last-place aversion" is particularly strong for conservatives and individuals with low income. Subgroup analyses by political orientation, gender, and income are presented in columns (3) - (8) of Table B.11.[15]

The most notable difference by political orientation is that the disadvantageous relative-pts-$ shock is significantly mood-diminishing for conservatives and has no significant mood-effect for liberal/progressives. Further, the mood-effects for conservatives and liberal/ progressives are marginally significantly different (-2.260 versus -0.106, p-value = 0.056).[16] In sum, an interesting asymmetry by political orientation is suggested: being "richer" may not affect conservatives and liberal/progressives differently, but being "poorer" may. This would be consistent with Kuziemko et al. (2014) and may be explained by conservatives' strong last-place aversion.

Subgroup analyses by gender reveal that the impact of the advantageous relative-pts-$, advantageous relative-pts, and disadvantageous relative-pts-$ shocks are significantly mood-diminishing for women but not men. Further, the mood-effect of the advantageous relative-pts shock is significantly greater in magnitude for women than men (-1.435 versus -0.296, p-value = 0.031). In sum, there is evidence that both being "richer" and being

[15]Because we divide the sample into subgroups, we use Bonferroni adjustments in these columns. As in Table B.10, Bonferroni adjustments are applied to alphas, not p-values; so the p-values reported in Table B.11 are unadjusted, and starring in the table corresponds to whether the raw p-value is less than the Bonferroni-adjusted alpha. Bonferroni-adjusted alphas are equal to the standard Type 1 error rate (e.g., 0.05) divided by two, the number of mutually exclusive subgroups the sample is divided into (e.g., male and female).

[16]Because the mood-effects of the disadvantageous relative-pts shocks are similar by political orientation, this means the mood-effects of the disadvantageous relative-$ shocks are different for conservatives and liberal/progressives, though not statistically significantly so at conventional levels (-2.101 versus 0.608, p-value = 0.107).

"poorer" are mood-diminishing for women but neither has mood-effects for men. This finding is consistent with Kamas and Preston (2015) and may be explained by women being more inequity-averse than men. There is also weak evidence of gender differences in the mood-effects of own-pts versus own-$ shocks: while the impact of the advantageous own-pts-$ shock is similar for men and women, the impact of the advantageous own-pts shock is marginally significantly more mood-improving for women than men (1.901 versus 0.915, p-value = 0.078).

Lastly, the only notable difference by income-subgroups is that the disadvantageous relative-pts-$ shock is significantly mood-diminishing for individuals with household income in the top-quartile and not for those in the bottom 3 quartiles.[17] This finding is consistent with Ifcher et al. (2018) and Alesina et al. (2004), but not Kuziemko et al. (2014).

3.5 Discussion

In an experiment designed to test for the RIE, we find no support for an interpretation that applies strictly to income; this is because we find that the impact on mood of being "richer" or "poorer" in experimental points is the same when points are worth money and when they are not. This cannot be explained by money and points impacting mood equally, as receiving more money than expected improves subjects' moods by significantly more than does receiving equivalently more points than expected. Further, we find only partial support for an interpretation of a generalized RE that applies to advantageous and disadvantageous relative-rewards shocks alike; this is because we find evidence of a worsening of mood from being "poorer" than others, but no evidence of an improvement in mood from being "richer" than others.

Our attempt to disentangle the RIE's income from non-income relativity concerns is

[17] The advantageous relative-pts shock does not significantly impact the mood of individuals with household income in the top quartile, but it is significantly mood-diminishing for those in the bottom 3 quartiles. While this may seem like an important difference, the similar magnitudes of the point-estimates (-0.776 versus -0.855, p-value = 0.909) suggest that it is likely due to variant sample sizes (n = 200 versus 646).

novel. Further, we do so with an extreme counterfactual: non-income relativity is manipulated using a non-monetized, cardinal point system.

What are the implications of our study for research that tests for the RIE using SWB and income data from surveys? As noted in the literature review, in such studies, the SWB-impact of non-relative factors like cost-of-living may be misattributed to others' income in the absence of proper controls. Our findings suggest that, even if a significant negative coefficient on others' income obtains with the inclusion of controls for non-relative factors like cost-of-living, the negative coefficient may not be attributable to concerns about how one's own income compares to others', but instead to other social-comparison concerns. Indeed, people may be biased toward using income because it is often "observed" as a number (e.g., a statistic) or as a material good (e.g., by proxy as conspicuous consumption), and because numbers and material income may be used as a proxy for ineffable factors in social comparisons that, if defined and concretized, would weaken relative income's relationship to SWB.

For example, suppose SWB's only determinant is feeling like a productive member of society—a latent variable for researchers.[18] Further suppose that though the feeling itself is observable to the individual having it, it is determined by a latent process that takes as inputs signals of others' productivity, own productivity, and socially held standards of productivity. Four cases emerge, depending on the availability of more direct productivity information: (1) If information on productivity is unavailable to both researchers and individuals, then the individual will use observed conspicuous consumption to infer others' income to, in turn, infer others' and own productivity (and/or use reported income data to infer productivity). In this case, RIE researchers will attribute traditional regression results to RIE. (2) If information on productivity is available to researchers but not individuals, individuals will continue to use conspicuous consumption and/or reported income data to infer productivity. Controlling for productivity will diminish the coefficient on others' income. If the coefficient is not nullified, the RIE will be supported, and the role of

[18]The example works just as well if we substitute "feeling like a productive member of society" with a possibly even more ineffable factor like "feeling high status."

productivity ruled out. (3) If information on productivity is available to individuals but not researchers, individuals will cease to use signals to infer productivity. If because of correlation between productivity and the input signals, the coefficient on others' income is still significant, researchers will attribute it to the RIE. (4) Only in the case that information on productivity is available to both researchers and individuals will regression analysis rule out the RIE and properly attribute SWB effects to productivity. Prior evidence regarding the RIE implicitly measures the impact of something akin to our relative-pts-$, rather than relative-$, shocks. The implication for RIE researchers depends on whether they are interested in distinguishing between relative income as a determinant of SWB and relative income as a signal-of-convenience for other determinants of SWB.[19]

In our experiment, of course, there is no scope for productivity, and subjects are informed, depending on treatment, either of a perfect correlation between points and money or a zero correlation between points and money. That we observe relative-pts shocks affecting mood suggests that numerical information, innately hierarchical as it is, is particularly salient in the latent process by which social information affects feelings. While points are explicitly non-monetized in our points treatment, their usual connotation as a reward together with the ease with which they, as numbers, convey relative standing may be enough to generate their mood-effects. This may have the same perceptual roots as money illusion, the observed tendency to understand prices in nominal rather than real terms. In related work, Butler (2016) conducts a laboratory experiment in which subjects are randomized into receiving high or low wages for the same task; they then perform the task and report their beliefs about their performance. Despite objectively equal task-performance and subjects' knowing their wages have been randomized, those assigned to high wages believe their performance to be significantly better than those assigned to low wages.

[19]Other experimental evidence of money's protean nature as a signal comes from the experimental psychology literature. Priming the concept of money (by, for example, showing subjects images of money or having them handle paper currency) increases subjects' self-sufficiency (Vohs, Mead, & Goode, 2006), tolerance for physical pain (Zhou, Vohs, & Baumeister, 2009), work and productivity (Mogilner, 2010); and decreases the savoring of experiences (Quoidbach, Dunn, Petrides, & Mikolajczak, 2010), willingness to volunteer or donate (Pfeffer & DeVoe, 2009; Chatterjee, Rose, & Sinha, 2013), socializing and happiness (Mogilner, 2010), and anxiety and fear of death (Zaleskiewicz, Gasiorowska, Kesebir, Luszczynska, & Pyszczynski, 2013).

Our research also has possible implications for policy makers, particularly those for whom citizens' SWB is an outcome of interest. First let's consider the policy implications of the RIE. The RIE poses a problem for policy makers seeking to increase SWB by pursuing policies that increase income across the distribution: the SWB gains from increases to individuals' own income would be at least partially offset by SWB losses from increases to others' income. The RIE does not advance the neoclassical economic understanding of the optimal mean-preserving distribution of income: as the variance of income increases, any SWB gains for those with higher-than-average income may be offset by SWB losses for those with lower-than-average income. Our results suggest zero SWB gains for those with higher-than-average income and SWB losses for those with lower-than-average income; this means that, *ceteris paribus*, more compressed and perhaps left-skewed income distributions would be favored.

Our approach, of course, begs the question of whether SWB-change is predictive of choice, and whether a revealed-preference approach would result in the same patterns we observe. In Benjamin, Heffetz, Kimball, and Rees-Jones (2014), graduating medical students' actual preference-rankings of residency programs are compared to their reported beliefs about various features of each program and their predictions of various aspects of their own experience (including their SWB) during and after each program. Predicted SWB is found to be the best, but not the only important, predictor of actual preference-rankings. While this study sheds some light on the relationship between SWB and revealed-preference, SWB is only considered prospectively and not contemporaneously. Further research is necessary to directly address whether our mood-change approach is compatible with a revealed-preference approach. If so, it could substitute for revealed-preference in other contexts, like identifying preferences for earned versus unearned income, or estimating parameters related to loss aversion, time-, risk-, or competitive-preferences. If not, it could potentially complement revealed-preference. For example, it may be that our approach, being emotionally based, reflects System-1 thinking while revealed-preference, being deliberative, reflects System-2 thinking, and that preferences may be best understood using a combination of

the two approaches (Kahneman, 2011).

Appendix A: Previous Approaches

Blanco et al. (2011) present the earliest attempt at testing the predictive power of F&S at the individual level. They employ decisions made by the responder in an ultimatum game to elicit α_i, and by the dictator in a modified dictator game to elicit β_i. They use the strategy method in both cases to calculate an approximate point estimate for each parameter based on the switching point from option A to option B for each individual (see table A.1).

In their ultimatum game, participants bargain over the allocation of £20. The proposer makes an offer, which is restricted to integer values. The responder states whether she accepts or rejects every possible offer. The switching point from accepting to rejecting, i.e. moving from option A to option B, indicates the responder i is indifferent between rejecting and accepting an offer $s_i \in [s'_i - 1, s'_i]$, s'_i being the lowest offer responder i is willing to accept, and $s'_i - 1$ being the highest offer i rejects. The indifference point revealed by the switching point signifies that $U_i(s_i, 20 - s_i) \equiv s_i - \alpha_i(20 - s_i - s_i) = 0$. Therefore, $\alpha_i = \frac{s_i}{2(10-s_i)}$, where $s_i = s'_i - 0.5$.

By a similar analysis, in the modified dictator game, Blanco et al. (2011) define $\beta_i = 1 - \frac{\widetilde{x}_i}{20}$, where $\widetilde{x}_i \in [x'_i - 1, x'_i]$. In this case, the indifference point is also approximated by the middle point between $x'_i - 1$ and x'_i, thus $\widetilde{x}_i = x'_i - 0.5$. Daruvala (2010) uses a similar approach in terms of determining the indifference between equality and an unequal payoff distribution. Unlike Blanco et al.'s, Daruvala's study is not in the context of a modified dictator game, and he studies groups of 11 individuals instead of pairs. Nonetheless, likewise Blanco et al. (2011), in Daruvala (2010) uses an environment where individuals may also be motivated by reciprocity. Yang et al. (2016) offer an alternative approach: instead of using standard games, they use a choice menu for each parameter (see table A.2). Individuals are asked to choose between to sets of allocations, either option A or option B, in each row of the choice menu. Intervals for α_i and β_i are inferred from the switching point in each menu. Similarly to the way parameters are estimated by Blanco et al. (2011), switching to

option B in row k of menu 1 means that:

$$\alpha_i \in \Big[\frac{Y_{Bk-1} - Y_{Ak-1}}{(O_{Bk-1} + Y_{Ak-1}) - (Y_{Bk-1} + O_{Ak-1})}, \frac{Y_{Bk} - Y_{Ak}}{(O_{Bk} + Y_{Ak}) - (Y_{Bk} + O_{Ak})}\Big],$$

where Y indicates Yours and O indicates Other's. Likewise, switching to option B in row k of menu 2 means that:

$$\beta_i \in \Big[\frac{Y_{Bk-1} - Y_{Ak-1}}{(Y_{Bk-1} + O_{Ak-1}) - (O_{Bk-1} + Y_{Ak-1})}, \frac{Y_{Bk} - Y_{Ak}}{(Y_{Bk} + O_{Ak}) - (O_{Bk} + Y_{Ak})}\Big]$$

Table A.1: Instrument used by Blanco et al. (2011)

(a) Ultimatum

Option A		Option B		α_i
Sender	Responder	Sender	Responder	
20	0	0	0	0.00
19	1	0	0	0.06
18	2	0	0	0.13
17	3	0	0	0.21
16	4	0	0	0.33
15	5	0	0	0.50
14	6	0	0	0.75
13	7	0	0	1.17
12	8	0	0	2.00
11	9	0	0	4.50
10	10	0	0	–
9	11	0	0	–
8	12	0	0	–
7	13	0	0	–
6	14	0	0	–
5	15	0	0	–
4	16	0	0	–
3	17	0	0	–
2	18	0	0	–
1	19	0	0	–
0	20	0	0	–

(b) Modified dictator

Option A		Option B		β_i
Sender	Responder	Sender	Responder	
20	0	0	0	1.00
20	0	1	1	0.95
20	0	2	2	0.90
20	0	3	3	0.85
20	0	4	4	0.80
20	0	5	5	0.75
20	0	6	6	0.70
20	0	7	7	0.65
20	0	8	8	0.60
20	0	9	9	0.55
20	0	10	10	0.50
20	0	11	11	0.45
20	0	12	12	0.40
20	0	13	13	0.35
20	0	14	14	0.30
20	0	15	15	0.25
20	0	16	16	0.20
20	0	17	17	0.15
20	0	18	18	0.10
20	0	19	19	0.05
20	0	20	20	0.00

Table A.2: Instrument used by Yang et al. (2016)

(a) Menu 1

Option A		Option B		α_i
Yours	Other's	Yours	Other's	
125	150	100	260	-0.19
115	150	100	260	-0.12
105	150	100	260	-0.04
95	150	100	260	0.05
85	150	100	260	0.16
75	150	100	260	0.29
65	150	100	260	0.47
55	150	100	260	0.69
45	150	100	260	1.00
35	150	100	260	1.44

(b) Menu 2

Option A		Option B		β_i
Yours	Other's	Yours	Other's	
185	90	170	50	-0.60
175	90	170	50	-0.14
165	90	170	50	0.11
155	90	170	50	0.27
145	90	170	50	0.38
135	90	170	50	0.47
125	90	170	50	0.53
115	90	170	50	0.58
105	90	170	50	0.62
95	90	170	50	0.65

Appendix B: Tables

Table B.1: Probit estimation (marginal effects)

DEPENDENT VARIABLE: Dummy=1 if i weakly adheres to the model	List 1	List 2	List 3
Female	0.031	0.055	0.056
	(0.113)	(0.050)	(0.094)
Age	-0.013	-0.003	-0.008
	(0.015)	(0.005)	(0.011)
Mother completed college	0.014	-0.001	0.059
	(0.125)	(0.050)	(0.091)
Father completed college	-0.008	-0.048	0.065
	(0.118)	(0.041)	(0.095)
Conservative political views	0.224	-0.903***	-0.045
	(0.158)	(0.056)	(0.150)
Progressive political views	0.124	-0.971***	-0.197
	(0.162)	(0.027)	(0.182)
Economist	-0.216	-0.302*	-0.129
	(0.193)	(0.160)	(0.158)
GPA	-0.026	0.029	0.156
	(0.143)	(0.053)	(0.120)
N	64	64	64

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: (1) Economist is coded as 1 for Economics and Business majors, including Accounting, Finance, Management and Marketing. (2) Regressions are estimated with a sample size of only 64, even thought 106 subjects participated in the experiment, because of missing values due to a lack of willingness of subjects to provide answers to certain questions.

Table B.2: Parameters' differences by demographic characteristics

	α		β	
	Average	p-value	Average	p-value
Gender	Females = 0.320 Males = 0.293	0.611	Females = 0.139 Males = 0.119	0.403
Age	18-20 years = 0.343 > 20 years = 0.278	0.126	18-20 years = 0.143 > 20 years = 0.119	0.246
Mother's education*	College = 0.299 Less than college = 0.317	0.126	College = 0.144 Less than college = 0.106	0.862
Father's education*	College = 0.315 Less than college = 0.293	0.702	College = 0.143 Less than college = 0.109	0.341
Economics major	Yes = 0.192 No = 0.318	0.052	Yes = 0.108 No = 0.131	0.421
GPA	$\geq 3.5 = 0.314$ $< 3.5 = 0.300$	0.726	$\geq 3.5 = 0.111$ $< 3.5 = 0.142$	0.964
Political views:				
Conservative	Yes = 0.318 No = 0.314	0.815	Yes = 0.116 No = 0.108	0.664
Moderate	Yes = 0.319 No = 0.311	0.842	Yes = 0.106 No = 0.113	0.884
Liberal	Yes = 0.308 No = 0.319	0.720	Yes = 0.111 No = 0.108	0.895

†Bonferroni-adjusted critical values correspond to *** p<0.001, ** p<0.005, * p<0.01

Note: * Parents' educational attainment is measured in terms of college attendance in this case: 'College' refers to having completed college or more, while 'Less than college' refers to attending college but not graduating or less.

† Critical values are adjusted to take into account multiple hypothesis testing, using Bonferroni correction. Critical values are divided into 9 hypothesis that were tested for each parameter: α and β.

Table B.3: Joint significance tests post Tobit estimation of equations (1.5) and (1.6)

		No rejection values for γ_0 and γ_1
$H_0 : \alpha_{i1} = \alpha_{i2} + \epsilon$	$\gamma_1 = 1,$	$\gamma_0 \in [-0.051, -0.023]^{\dagger}$ $\gamma_0 \in [-0.061, -0.013]^{\dagger\dagger}$
$H_0 : \alpha_{i1} = \alpha_{i3} + \epsilon$	$\gamma_1 = 1,$	$\gamma_0 \in [-0.018, +0.046]^{\dagger}$ $\gamma_0 \in [-0.024, +0.052]^{\dagger\dagger}$
$H_0 : \alpha_{i2} = \alpha_{i3} + \epsilon$	$\gamma_1 = 1,$	$\gamma_0 \in [+0.015, +0.070]^{\dagger}$ $\gamma_0 \in [+0.006, +0.078]^{\dagger\dagger}$
$H_0 : \beta_{i1} = \beta_{i2} + \epsilon$	$\gamma_1 = 1,$	$\gamma_0 \in [-0.175, -0.125]^{\dagger}$ $\gamma_0 \in [-0.192, -0.107]^{\dagger\dagger}$
$H_0 : \beta_{i1} = \beta_{i3} + \epsilon$	$\gamma_1 = 1,$	$\gamma_0 \in [-0.138, -0.038]^{\dagger}$ $\gamma_0 \in [-0.148, -0.029]^{\dagger\dagger}$
$H_0 : \beta_{i2} = \beta_{i3} + \epsilon$	$\gamma_1 = 1,$	$\gamma_0 \in [-0.034, +0.063]^{\dagger}$ $\gamma_0 \in [-0.041, +0.070]^{\dagger\dagger}$

†† p ≥ 0.05, † p ≥ 0.1

Note: This table reports significance in a reversed way. We are showing the range of values for γ_0 such that each null hypothesis is not rejected at 5% †† and 10% † significance levels.

Table B.4: Scoring results by individual: Part 1

Subject	F&S	F&S Guilt	F&S Envy	F&S no guilt, no envy	RIE	Capped RIE	Equality aversion	Efficiency concerns	Altruist
1	1	9	11	**19**	1	9	0	0	10
2	9	5	**15**	11	7	3	0	2	8
3	11	1	**16**	6	13	3	3	1	6
4	5	2	**10**	7	**10**	7	8	3	8
5	4	8	12	**16**	3	7	0	1	9
6	9	1	**19**	11	9	1	0	0	10
7	7	**13**	3	9	0	6	4	11	7
8	0	10	10	**20**	0	10	0	0	10
9	11	1	11	1	**18**	8	8	1	1
10	**18**	8	12	2	10	0	0	8	2
11	**17**	7	12	2	11	1	1	7	2
12	2	10	10	**18**	1	9	0	1	9
13	**11**	**11**	8	8	6	6	1	6	3
14	7	**15**	5	13	1	9	0	6	4
15	4	5	14	**15**	4	5	1	1	11
16	14	**16**	4	6	4	6	0	10	0
17	1	9	11	**19**	1	9	0	0	10
18	12	2	**16**	6	12	2	2	2	6
19	3	9	11	**17**	2	8	0	1	9
20	6	7	11	**12**	5	6	2	3	8
21	9	1	**19**	11	9	1	0	0	10
22	**20**	10	10	0	10	0	0	10	0
23	**15**	13	7	5	6	4	0	9	1
24	**14**	6	12	4	11	3	2	5	3
25	6	4	**16**	14	6	4	0	0	10
26	5	12	6	**13**	2	9	2	5	6
27	0	10	10	**20**	0	10	0	0	10
28	**13**	6	**13**	6	8	1	1	6	6
29	6	10	10	**14**	3	7	0	3	7
30	2	11	8	**17**	0	9	1	3	9
31	**11**	10	6	5	5	4	4	10	5
32	1	9	11	**19**	1	9	0	0	10
33	0	10	10	**20**	0	10	0	0	10
34	9	3	**17**	11	8	2	0	1	9
35	6	**14**	5	13	2	10	1	5	4
36	4	5	9	**10**	9	**10**	6	1	6
37	4	6	14	**16**	4	6	0	0	10
38	**18**	8	11	1	11	1	1	8	1
39	1	9	11	**19**	1	9	0	0	10
40	4	12	5	**13**	1	9	3	6	7
41	5	2	**15**	12	5	2	3	3	13
42	6	4	**15**	13	7	5	1	0	9
43	10	0	**20**	10	10	0	0	0	10
44	8	3	9	4	**11**	6	8	5	6
45	5	5	**15**	**15**	5	5	0	0	10
46	1	9	11	**19**	1	9	0	0	10
47	**17**	9	11	3	9	1	0	8	2
48	7	2	**17**	12	7	2	1	1	11
49	2	8	11	**17**	3	9	1	0	9
50	9	**17**	3	11	1	9	0	8	2
51	9	1	**18**	10	10	2	1	0	9
52	1	4	10	13	2	5	6	5	**14**

Table B.5: Scoring results by individual: Part 2

Subject	F&S	F&S Guilt	F&S Envy	F&S no guilt, no envy	RIE	Capped RIE	Equality aversion	Efficiency concerns	Altruist
53	**13**	7	**13**	7	8	2	0	5	5
54	6	4	**16**	14	6	4	0	0	10
55	6	4	**12**	10	10	8	4	0	6
56	10	0	**20**	10	10	0	0	0	10
57	**11**	9	**11**	9	6	4	0	5	5
58	9	1	**19**	11	9	1	0	0	10
59	10	0	**20**	10	10	0	0	0	10
60	1	9	11	**19**	1	9	0	0	10
61	8	4	**14**	10	9	5	2	1	7
62	2	9	10	**17**	1	8	1	2	10
63	5	13	7	**15**	1	9	0	4	6
64	**15**	11	9	5	7	3	0	8	2
65	11	2	**15**	6	11	2	3	3	7
66	9	1	12	4	**16**	8	7	0	3
67	0	10	8	**18**	2	12	2	0	8
68	10	**20**	0	10	0	10	0	10	0
69	10	0	**15**	5	**15**	5	5	0	5
70	0	10	10	**20**	0	10	0	0	10
71	**18**	8	12	2	10	0	0	8	2
72	3	3	10	10	6	6	7	4	**11**
73	8	4	**15**	11	8	4	1	1	8
74	2	8	12	**18**	2	8	0	0	10
75	5	5	**13**	**13**	7	7	2	0	8
76	8	2	**13**	7	**13**	7	5	0	5
77	8	2	**18**	12	8	2	0	0	10
78	5	4	**15**	14	5	4	1	1	11
79	9	**11**	5	7	5	7	4	8	4
80	**20**	10	10	0	10	0	0	10	0
81	**18**	8	12	2	10	0	0	8	2
82	6	4	**15**	13	7	5	1	0	9
83	**14**	6	**14**	6	9	1	0	5	5
84	3	7	13	**17**	3	7	0	0	10
85	13	5	**15**	7	9	1	0	4	6
86	4	14	6	**16**	0	10	0	4	6
87	**20**	10	10	0	10	0	0	10	0
88	6	4	**15**	13	7	5	1	0	9
89	3	8	7	**12**	6	11	5	2	6
90	7	3	**17**	13	7	3	0	0	10
91	1	11	9	**19**	0	10	0	1	9
92	3	7	13	**17**	3	7	0	0	10
93	**20**	10	10	0	10	0	0	10	0
94	**13**	5	12	4	12	4	3	4	3
95	10	0	**19**	9	11	1	1	0	9
96	2	8	12	**18**	2	8	0	0	10
97	0	10	9	**19**	1	11	1	0	9
98	10	0	**20**	10	10	0	0	0	10
99	9	0	**19**	10	9	0	1	1	11
100	8	2	**18**	12	8	2	0	0	10
101	5	11	9	**15**	2	8	0	3	7
102	**9**	6	**9**	6	7	4	5	7	7
103	13	3	**16**	6	11	1	1	3	6
104	14	4	**15**	5	11	1	1	4	5
105	**17**	9	11	3	9	1	0	8	2
106	7	3	7	3	**17**	13	10	0	0

Table B.6: 4X2 treatment-design, with own, Participant X's, and relative rewards by cell

Cell	pts-$-Treatment			pts-Treatment		
	Self	Participant X	Relative rewards	Self	Participant X	Relative rewards
LL	$2	$2	0	$6	$6	0
	2 pts	2 pts	0	2 pts	2 pts	0
LH	$2	$10	-$8	$6	$6	0
	2 pts	10 pts	-8 pts	2 pts	10 pts	-8 pts
HL	$10	$2	+$8	$6	$6	0
	10 pts	2 pts	+8 pts	10 pts	2 pts	+8 pts
HH	$10	$10	0	$6	$6	0
	10 pts	10 pts	0	10 pts	10 pts	0

Note: As the instructions disclosed all possible revelations, the subject's expected pts- and $-rewards are 6 pts and $6 for both herself and Participant X for all cells. Relative rewards are own minus Participant X's rewards.

Table B.7: Relative- and own-rewards shocks and formulae used for estimation

A. Shocks	B. Difference-of-means estimators	C. Regression estimators
Advantageous relative-pts-$	$\Delta\text{MSF}_{\text{HL,pts\$}}$ - $\Delta\text{MSF}_{\text{HH,pts\$}}$	$-(\beta_4+\beta_6+\beta_7)=(\beta_0+\beta_2+\beta_3)-(\beta_0+\beta_2+\beta_3+\beta_4+\beta_6+\beta_7)$
Advantageous relative-pts	$\Delta\text{MSF}_{\text{HL,pts}}$ - $\Delta\text{MSF}_{\text{HH,pts}}$	$-(\beta_4+\beta_7)=(\beta_0+\beta_3)-(\beta_0+\beta_3+\beta_4+\beta_7)$
Advantageous relative-$	($\Delta\text{MSF}_{\text{HL,pts\$}}$ - $\Delta\text{MSF}_{\text{HH,pts\$}}$)-($\Delta\text{MSF}_{\text{HL,pts}}$ - $\Delta\text{MSF}_{\text{HH,pts}}$)	$-\beta_6=-(\beta_4+\beta_6+\beta_7)+(\beta_4+\beta_7)$
Disadvantageous relative-pts-$	$\Delta\text{MSF}_{\text{LH,pts\$}}$ - $\Delta\text{MSF}_{\text{LL,pts\$}}$	$\beta_4+\beta_5=(\beta_0+\beta_1+\beta_4+\beta_5)-(\beta_0+\beta_1)$
Disadvantageous relative-pts	$\Delta\text{MSF}_{\text{LH,pts}}$ - $\Delta\text{MSF}_{\text{LL,pts}}$	$\beta_4=(\beta_0+\beta_4)-\beta_0$
Disadvantageous relative-$	($\Delta\text{MSF}_{\text{LH,pts\$}}$ - $\Delta\text{MSF}_{\text{LL,pts\$}}$)-($\Delta\text{MSF}_{\text{LH,pts}}$ - $\Delta\text{MSF}_{\text{LL,pts}}$)	$\beta_5=(\beta_4+\beta_5)-\beta_4$
Advantageous own-pts-$	$\Delta\text{MSF}_{\text{HH,pts\$}}$ - $\Delta\text{MSF}_{\text{LL,pts\$}}$	$-\beta_1+\beta_2+\beta_3+\beta_4+\beta_6+\beta_7=(\beta_0+\beta_2+\beta_3+\beta_4+\beta_6+\beta_7)-(\beta_0+\beta_1)$
Advantageous own-pts	$\Delta\text{MSF}_{\text{HH,pts}}$ - $\Delta\text{MSF}_{\text{LL,pts}}$	$\beta_3+\beta_4+\beta_7=(\beta_0+\beta_3+\beta_4+\beta_7)-\beta_0$
Advantageous own-$	($\Delta\text{MSF}_{\text{HH,pts\$}}$ - $\Delta\text{MSF}_{\text{LL,pts\$}}$)-($\Delta\text{MSF}_{\text{HH,pts}}$ - $\Delta\text{MSF}_{\text{LL,pts}}$)	$-\beta_1+\beta_2+\beta_6=(-\beta_1+\beta_2+\beta_3+\beta_4+\beta_6+\beta_7)-(\beta_3+\beta_4+\beta_7)$

Table B.8: Demographic characteristics, n = 860

	Mean
Gender	
Female	0.45
Male	0.54
Other / prefer not answer	0.01
Age	36.48
Are you a citizen or permanent resident of the United States?	
Yes	1.00
No	0.00
Prefer not to answer	0.00
Rate your English	
Native	0.98
Fluent	0.02
Proficient	0.00
What race / ethnicity do you identify yourself as?	
American Indian and Alaska native	0.00
Asian	0.08
Black or African	0.07
Hispanic	0.05
White	0.77
Other / prefer not to answer	0.02
What religion do you consider yourself	
Atheist / agnostic	0.47
Buddhist	0.02
Christian	0.42
Hindu	0.00
Jewish	0.02
Muslim	0.00
Other / Prefer not to answer	0.07
How would you characterize your political views?	
Conservative	0.20
Liberal / progressive	0.50
Moderate	0.28
Prefer not to answer	0.02
what is the total (gross) income last year of your household?	
$0-$25,000	0.18
$25,000-$50,000	0.30
$50,000-$75,000	0.24
$75,000-$100,000	0.15
$100,000-$125,000	0.05
$125,000-$150,000	0.03
$150,000+	0.04
Prefer not to answer	0.02
What is your highest level of education?	
Some high school	0.00
Completed high school	0.10
Some college	0.28
Completed college	0.44
Some grad / professional school	0.04
Completed grad / professional school	0.14
Prefer not to answer	0.00
Please indicate your employment status:	
Employed, full-time	0.67
Employed, part-time	0.15
Not employed, looking for work	0.06
Not employed, not looking for work	0.06
Retired	0.02
Student	0.03
Prefer not to answer	0.02

Table B.9: Pre- and post-revelation MSF-scores and ΔMSF by cell

Cell	MSF-score	pts-$-Treatment	pts-Treatment
LL	pre-revelation	15.081	15.495
		(3.451)	(2.950)
	post-revelation	13.970	15.495
		(3.947)	(3.340)
	change	-1.111	0.000
		(2.740)	(1.788)
		p=0.0009	p=0.9417
	observations	99	103
LH	pre-revelation	14.922	15.055
		(3.389)	(3.410)
	post-revelation	12.767	14.312
		(4.225)	(3.736)
	change	-2.155	-0.743
		(3.247)	(2.605)
		p=0.0000	p=0.0036
	observations	103	109
HL	pre-revelation	15.398	16.056
		(3.109)	(2.936)
	post-revelation	17.293	16.539
		(2.983)	(2.735)
	change	1.894	0.483
		(2.472)	(1.439)
		p=0.0000	p=0.0051
	observations	123	89
HH	pre-revelation	15.030	15.139
		(3.481)	(3.516)
	post-revelation	17.286	16.337
		(2.933)	(3.226)
	change	2.256	1.198
		(2.338)	(1.709)
		p=0.0000	p=0.0000
	observations	133	101

Note: Standard deviations are in parentheses. p-values are from Wilcoxon signed-rank tests of the equality of the distributions from which pre- and post-revelation MSF-scores are drawn.

Table B.10: Estimated impacts of relative- and own-rewards shocks on ΔMSF

Shocks	Formulae	
Adv own-pts-$	$\Delta MSF_{HH,pts\$}$ - $\Delta MSF_{LL,pts\$}$	3.367***
		(0.342)
		p=0.0000
Adv own-pts	$\Delta MSF_{HH,pts}$ - $\Delta MSF_{LL,pts}$	1.198***
		(0.245)
		p=0.0000
Adv own-$	($\Delta MSF_{HH,pts\$}$ - $\Delta MSF_{LL,pts\$}$)-($\Delta MSF_{HH,pts}$ - $\Delta MSF_{LL,pts}$)	2.169***
		(0.421)
		p=0.0000
Adv rel-pts-$	$\Delta MSF_{HL,pts\$}$ - $\Delta MSF_{HH,pts\$}$	-0.361
		(0.302)
		p=0.3256
Adv rel-pts	$\Delta MSF_{HL,pts}$ - $\Delta MSF_{HH,pts}$	-0.715***
		(0.228)
		p=0.0021
Adv rel-$	($\Delta MSF_{HL,pts\$}$ - $\Delta MSF_{HH,pts\$}$)-($\Delta MSF_{HL,pts}$ - $\Delta MSF_{HH,pts}$)	0.354
		(0.378)
		p=0.5775
Disadv rel-pts-$	$\Delta MSF_{LH,pts\$}$ - $\Delta MSF_{LL,pts\$}$	-1.044**
		(0.422)
		p=0.0095
Disadv rel-pts	$\Delta MSF_{LH,pts}$ - $\Delta MSF_{LL,pts}$	-0.743*
		(0.305)
		p=0.0208
Disadv rel-$	($\Delta MSF_{LH,pts\$}$ - $\Delta MSF_{LL,pts\$}$)-($\Delta MSF_{LH,pts}$ - $\Delta MSF_{LL,pts}$)	-0.301
		(0.521)
		p=0.5618

Note: Standard deviations are in parentheses. p-values are from Mann-Whitney tests; for example, the first row tests the equality of distributions from which $\Delta MSF_{HH,pts\$}$ and $\Delta MSF_{LL,pts\$}$ are drawn. Bonferroni adjusted alphas are used for starring so *, **, *** represent p-values < 0.0333, 0.0167, and 0.0033, respectively (representing standard Type 1 error rates of 0.10, 0.05, and 0.01 divided by the number of tests in a family; the advantageous-own-rewards family has three tests (first three rows); the advantageous-relative-rewards family has three tests (next three rows); and the disadvantageous-relative-rewards family has three tests (last three rows).

Table B.11: Regression estimates of the impacts of relative- and own-rewards shocks on ΔMSF

Shocks	All (1)	All (2)	Conservative (3)	Liberal/ Progressive (4)	Female (5)	Male (6)	Income in top quartile (7)	Income in bottom 3 quartiles (8)
Panel A: Impacts of advantageous relative-rewards shocks								
Adv rel-pts-$	-0.361	-0.431	-0.780	-0.554	-0.991*	-0.076	-0.369	-0.311
	(0.302)	(0.313)	(0.671)	(0.496)	(0.490)	(0.449)	(0.673)	(0.346)
	p=0.231	p=0.170	p=0.247	p=0.265	p=0.044	p=0.866	p=0.584	p=0.369
			p-value of equality=0.787		p-value of equality=0.169		p-value of equality=0.939	
Adv rel-pts	-0.715***	-0.800***	-0.953	-0.463	-1.435***	-0.296	-0.776	-0.855***
	(0.228)	(0.242)	(0.553)	(0.339)	(0.437)	(0.295)	(0.631)	(0.285)
	p=0.002	p=0.001	p=0.087	p=0.173	p=0.001	p=0.316	p=0.221	p=0.003
			p-value of equality=0.450		p-value of equality=0.031		p-value of equality=0.909	
Adv rel-$	0.354	0.370	0.173	-0.091	0.445	0.220	0.406	0.544
	(0.378)	(0.395)	(0.841)	(0.602)	(0.659)	(0.549)	(0.943)	(0.448)
	p=0.350	p=0.349	p=0.837	p=0.880	p=0.500	p=0.689	p=0.667	p=0.225
			p-value of equality=0.798		p-value of equality=0.793		p-value of equality=0.895	
Panel B: Impacts of disadvantageous relative-rewards shocks								
Disadv rel-pts-$	-1.044**	-1.026**	-2.620*	-0.106	-1.748**	-0.293	-1.975*	-0.788
	(0.422)	(0.432)	(1.176)	(0.590)	(0.664)	(0.598)	(0.925)	(0.503)
	p=0.014	p=0.018	p=0.028	p=0.858	p=0.009	p=0.624	p=0.034	p=0.0118
			p-value of equality=0.056		p-value of equality=0.104		p-value of equality=0.260	
Disadv rel-pts	-0.743**	-0.665**	-0.518	-0.714	-0.686	-0.500	-0.708	-0.520
	(0.305)	(0.318)	(0.538)	(0.405)	(0.511)	(0.413)	(0.607)	(0.399)
	p=0.015	p=0.037	p=0.337	p=0.078	p=0.180	p=0.227	p=0.245	p=0.193
			p-value of equality=0.772		p-value of equality=0.777		p-value of equality=0.795	
Disadv rel-$	-0.301	-0.361	-2.101	0.608	-1.061	0.207	-1.267	-0.268
	(0.521)	(0.537)	(1.294)	(0.718)	(0.822)	(0.702)	(1.120)	(0.647)
	p=0.563	p=0.501	p=0.107	p=0.397	p=0.198	p=0.769	p=0.260	p=0.679
			p-value of equality=0.067		p-value of equality=0.241		p-value of equality=0.440	
Panel C: Impacts of advantageous own-rewards shocks								
Adv own-pts-$	3.367***	3.392***	2.934***	4.136***	3.136***	3.625***	2.820***	3.548***
	(0.342)	(0.353)	(0.660)	(0.562)	(0.484)	(0.542)	(0.711)	(0.415)
	p=0.000	p=0.000	p=0.000	p=0.000	p=0.000	p=0.000	p=0.000	p=0.000
			p-value of equality=0.166		p-value of equality=0.502		p-value of equality=0.376	
Adv own-pts	1.198***	1.294***	1.208**	1.167***	1.901***	0.915**	0.898	1.369***
	(0.245)	(0.256)	(0.508)	(0.350)	(0.445)	(0.337)	(0.543)	(0.327)
	p=0.000	p=0.000	p=0.019	p=0.001	p=0.000	p=0.007	p=0.100	p=0.000
			p-value of equality=0.947		p-value of equality=0.078		p-value of equality=0.457	
Adv own-$	2.169***	2.098***	1.726*	2.969***	1.236	2.710***	1.922*	2.179***
	(0.421)	(0.433)	(0.845)	(0.643)	(0.650)	(0.637)	(0.886)	(0.529)
	p=0.000	p=0.000	p=0.043	p=0.000	p=0.058	p=0.000	p=0.031	p=0.000
			p-value of equality=0.242		p-value of equality=0.105		p-value of equality=0.803	
Controls included	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
R squared	0.278	0.311	0.477	0.347	0.360	0.326	0.423	0.310
Observations	860	860	175	434	391	463	200	646

Note: Robust standard errors are reported in parentheses. In columns (1) & (2), *, **, *** indicate p-values < 0.10, 0.05, and 0.01, respectively. Bonferroni adjusted alphas are used for starring in columns (3)-(8) to account for the two subgroups in columns (3)&(4), (5)&(6), and (7)&(8), respectively; so in those columns *, **, *** indicate p-values < 0.05, 0.025, and 0.005. p-values from tests of equality of coefficients across columns (3)&(4), (5)&(6), and (7)&(8) are presented below the corresponding coefficients.

Table B.12: Regression estimates of the impacts of relative- and own-rewards shocks on ΔMSF when including dropped subjects and when including pilot-study subjects, and on post-revelation 10-item BPNSS score

Shocks	Include subjects from pilot study (1)	Include subjects from pilot study (2)	Include dropped subjects (3)	Include dropped subjects (4)	Post-revelation BPNSS score (10 items) (5)	Post-revelation BPNSS score (10 items) (6)
Panel A: Impacts of advantageous relative-rewards shocks						
Adv rel-pts-$	-0.400	-0.429	-0.118	-0.118	0.092	0.114
	(0.269)	(0.313)	(0.278)	(0.284)	(0.124)	(0.127)
Adv rel-pts	-0.595***	-0.801***	-0.622***	-0.677***	0.115	0.118
	(0.200)	(0.242)	(0.235)	(0.251)	(0.131)	(0.124)
Adv rel-$	0.195	0.371	0.504	0.559	0.-0.023	-0.004
	(0.335)	(0.395)	(0.364)	(0.375)	(0.180)	(0.176)
Panel B: Impacts of disadvantageous relative-rewards shocks						
Disadv rel-pts-$	-1.019***	-1.026**	-0.990**	-0.939**	-0.152	-0.186
	(0.388)	(0.432)	(0.396)	(0.400)	(0.158)	(0.154)
Disadv rel-pts	-0.497*	-0.663**	-0.631**	-0.484	-0.066	-0.085
	(0.256)	(0.318)	(0.291)	(0.300)	(0.148)	(0.148)
Disadv rel-$	-0.522	-0.363	-0.359	-0.454	-0.087	-0.101
	(0.465)	(0.537)	(0.491)	(0.500)	(0.216)	(0.212)
Panel C: Impacts of advantageous own-rewards shocks						
Adv own-pts-$	3.187***	3.392***	3.100***	3.137***	-0.015	-0.048
	(0.315)	(0.353)	(0.314)	(0.321)	(0.149)	(0.148)
Adv own-pts	1.104***	1.297***	1.285***	1.366***	0.182	0.201
	(0.213)	(0.257)	(0.236)	(0.248)	(0.141)	(0.138)
Adv own-$	2.083***	2.096***	1.814***	1.771***	-0.197	-0.248
	(0.380)	(0.434)	(0.393)	(0.400)	(0.205)	(0.202)
Controls included	No	Yes	No	Yes	No	Yes
Observations	1,054	1,054	996	996	860	860

Note:Robust standard errors are reported in parentheses. *, **, *** indicate p-values < 0.10, 0.05, and 0.01, respectively. Controls include gender, age, race, religion, political leaning, household income, education, and employment status.

Appendix C: Figures

TASK: List 1

In this task, you must indicate your level of satisfaction with each allocation. Note that in every allocation you get $20, and what changes from one allocation to the next is what the person you are paired with gets.

How satisfied would you be?

	Extremely Dissatisfied 1	Moderately Dissatisfied 2	Slightly Dissatisfied 3	Neutral 4	Slightly Satisfied 5	Moderately Satisfied 6	Extremely Satisfied 7
If you get **$20**, and the other person gets **$10**							
If you get **$20**, and the other person gets **$11**							
If you get **$20**, and the other person gets **$12**							
If you get **$20**, and the other person gets **$13**							
If you get **$20**, and the other person gets **$14**							
If you get **$20**, and the other person gets **$15**							
If you get **$20**, and the other person gets **$16**							
If you get **$20**, and the other person gets **$17**							
If you get **$20**, and the other person gets **$18**							
If you get **$20**, and the other person gets **$19**							
If you get **$20**, and the other person gets **$20**							
If you get **$20**, and the other person gets **$21**							
If you get **$20**, and the other person gets **$22**							
If you get **$20**, and the other person gets **$23**							
If you get **$20**, and the other person gets **$24**							
If you get **$20**, and the other person gets **$25**							
If you get **$20**, and the other person gets **$26**							
If you get **$20**, and the other person gets **$27**							
If you get **$20**, and the other person gets **$28**							
If you get **$20**, and the other person gets **$29**							
If you get **$20**, and the other person gets **$30**							

>>

Figure C.1: List 1

Note: In the experiment participants receive instructions before completing Lists 1, 2 and 3. These instructions apply to all three lists and explain the payment procedure.

TASK: List 2

In this task, you must indicate your level of satisfaction with each allocation. Note that in every allocation you get $10, and what changes from one allocation to the next is what the person you are paired with gets.

How satisfied would you be?

	Extremely Dissatisfied 1	Moderately Dissatisfied 2	Slightly Dissatisfied 3	Neutral 4	Slightly Satisfied 5	Moderately Satisfied 6	Extremely Satisfied 7
If you get **$10** and the other person gets **$0**							
If you get **$10** and the other person gets **$1**							
If you get **$10** and the other person gets **$2**							
If you get **$10** and the other person gets **$3**							
If you get **$10** and the other person gets **$4**							
If you get **$10** and the other person gets **$5**							
If you get **$10** and the other person gets **$6**							
If you get **$10** and the other person gets **$7**							
If you get **$10** and the other person gets **$8**							
If you get **$10** and the other person gets **$9**							
If you get **$10** and the other person gets **$10**							
If you get **$10** and the other person gets **$11**							
If you get **$10** and the other person gets **$12**							
If you get **$10** and the other person gets **$13**							
If you get **$10** and the other person gets **$14**							
If you get **$10** and the other person gets **$15**							
If you get **$10** and the other person gets **$16**							
If you get **$10** and the other person gets **$17**							
If you get **$10** and the other person gets **$18**							
If you get **$10** and the other person gets **$19**							
If you get **$10** and the other person gets **$20**							

>>

Figure C.2: List 2

Note: In the experiment participants receive instructions before completing Lists 1, 2 and 3. These instructions apply to all three lists and explain the payment procedure.

TASK: List 3

In this task, you must indicate your level of satisfaction with each allocation. Note that in every allocation you get $20, and what changes from one allocation to the next is what the person you are paired with gets.

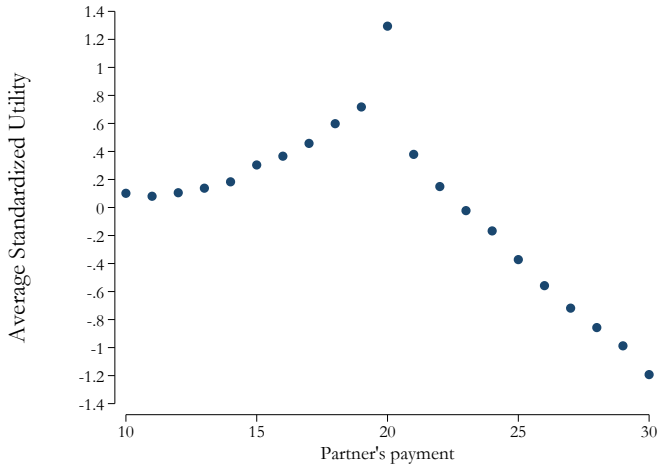
How satisfied would you be?

	Extremely Dissatisfied 1	Moderately Dissatisfied 2	Slightly Dissatisfied 3	Neutral 4	Slightly Satisfied 5	Moderately Satisfied 6	Extremely Satisfied 7
If you get **$20** and the other person gets **$1**							
If you get **$20** and the other person gets **$2**							
If you get **$20** and the other person gets **$3**							
If you get **$20** and the other person gets **$4**							
If you get **$20** and the other person gets **$5**							
If you get **$20** and the other person gets **$6**							
If you get **$20** and the other person gets **$7**							
If you get **$20** and the other person gets **$8**							
If you get **$20** and the other person gets **$9**							
If you get **$20** and the other person gets **$10**							
If you get **$20** and the other person gets **$11**							
If you get **$20** and the other person gets **$12**							
If you get **$20** and the other person gets **$13**							
If you get **$20** and the other person gets **$14**							
If you get **$20** and the other person gets **$15**							
If you get **$20** and the other person gets **$16**							
If you get **$20** and the other person gets **$17**							
If you get **$20** and the other person gets **$18**							
If you get **$20** and the other person gets **$19**							
If you get **$20** and the other person gets **$20**							
If you get **$20** and the other person gets **$21**							
If you get **$20** and the other person gets **$22**							
If you get **$20** and the other person gets **$23**							
If you get **$20** and the other person gets **$24**							
If you get **$20** and the other person gets **$25**							
If you get **$20** and the other person gets **$26**							
If you get **$20** and the other person gets **$27**							
If you get **$20** and the other person gets **$28**							
If you get **$20** and the other person gets **$29**							
If you get **$20** and the other person gets **$30**							
If you get **$20** and the other person gets **$31**							
If you get **$20** and the other person gets **$32**							
If you get **$20** and the other person gets **$33**							
If you get **$20** and the other person gets **$34**							
If you get **$20** and the other person gets **$35**							
If you get **$20** and the other person gets **$36**							
If you get **$20** and the other person gets **$37**							
If you get **$20** and the other person gets **$38**							
If you get **$20** and the other person gets **$39**							

>>

Figure C.3: List 3

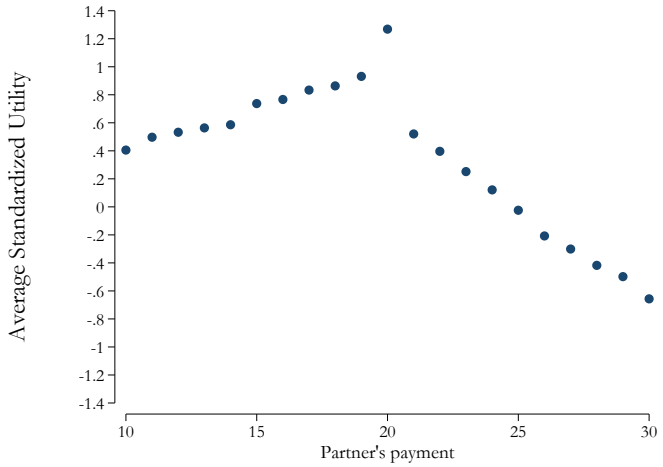
Note: In the experiment participants receive instructions before completing Lists 1, 2 and 3. These instructions apply to all three lists and explain the payment procedure.



N=106, $\alpha = 0.20$, $\beta = 0.09$

Figure C.4: Aggregate utility List 2

Note: Standardized utility is measured as $\frac{(U_{ik}-\overline{U_i})}{sd(U_i)}$, where U_{ik} is the utility expressed by individual i with regard to allocation k.



N=106, $\alpha = 0.16$, $\beta = 0.07$

Figure C.5: Aggregate utility List 3

Note: Standardized utility is measured as $\frac{(U_{ik}-\overline{U_i})}{sd(U_i)}$, where U_{ik} is the utility expressed by individual i with regard to allocation k.

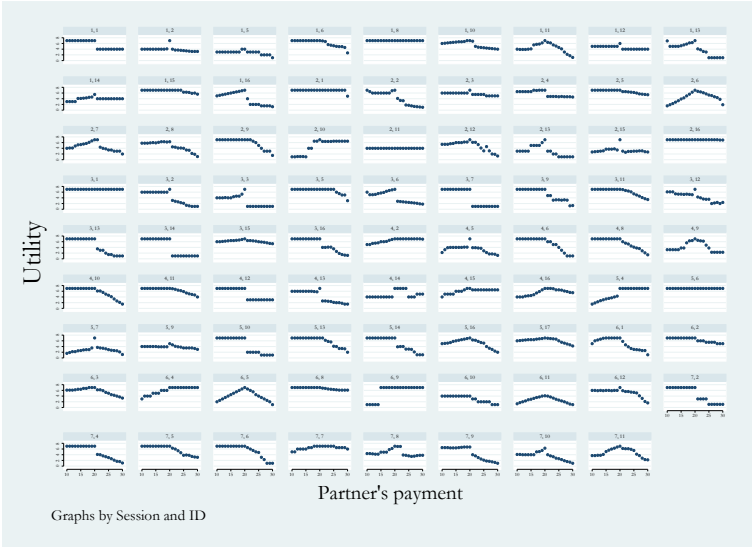


Figure C.6: Weak adherents to Fehr and Schmidt's model

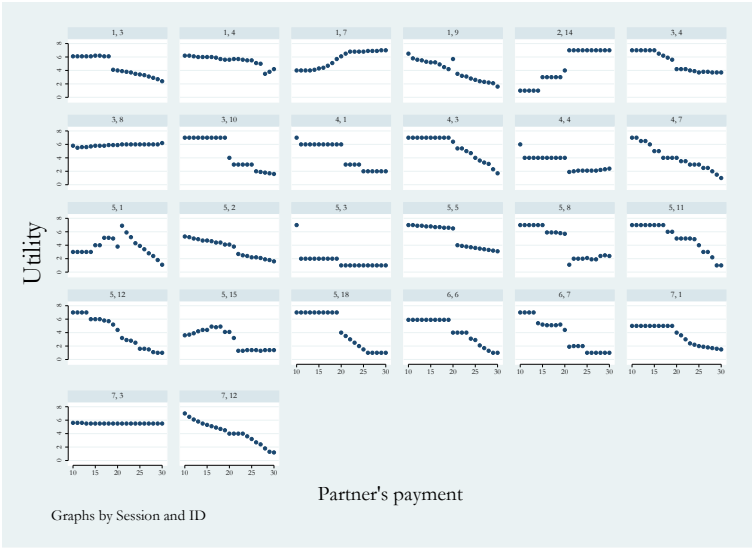


Figure C.7: Non-adherents to Fehr and Schmidt's model

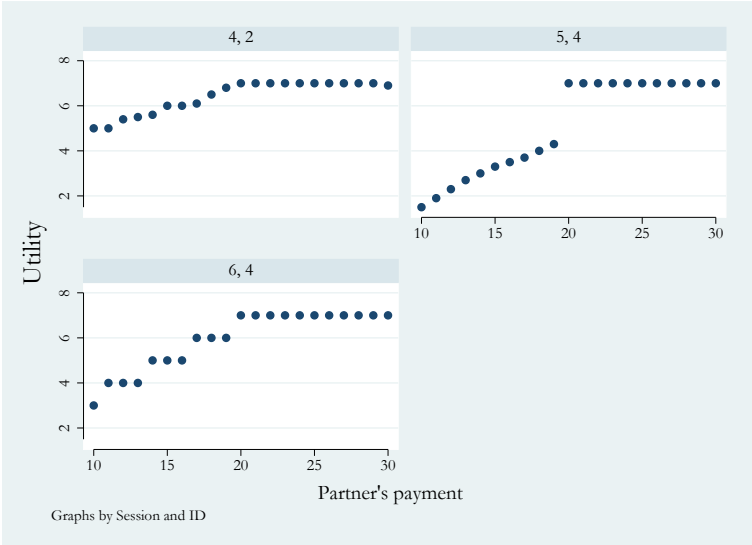


Figure C.8: Individuals displaying maximin preferences

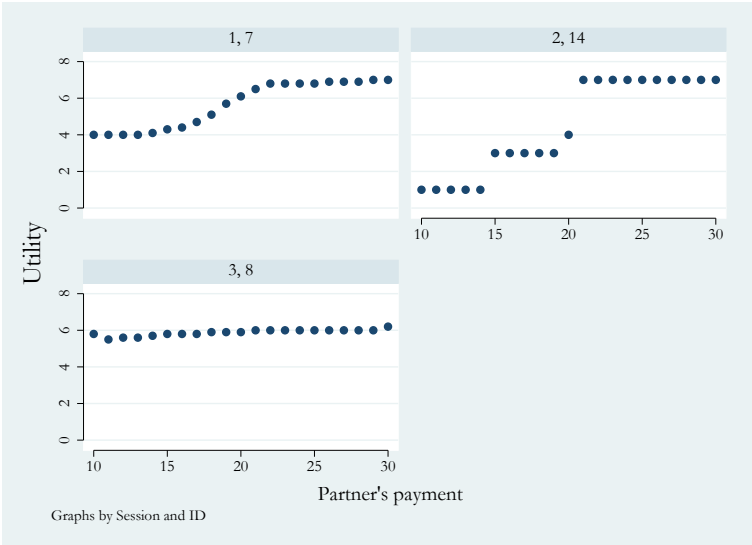


Figure C.9: Individuals displaying efficiency concerns

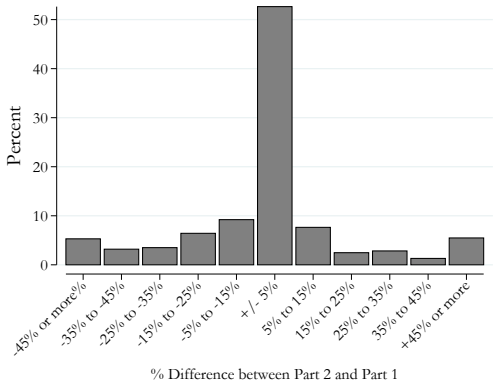


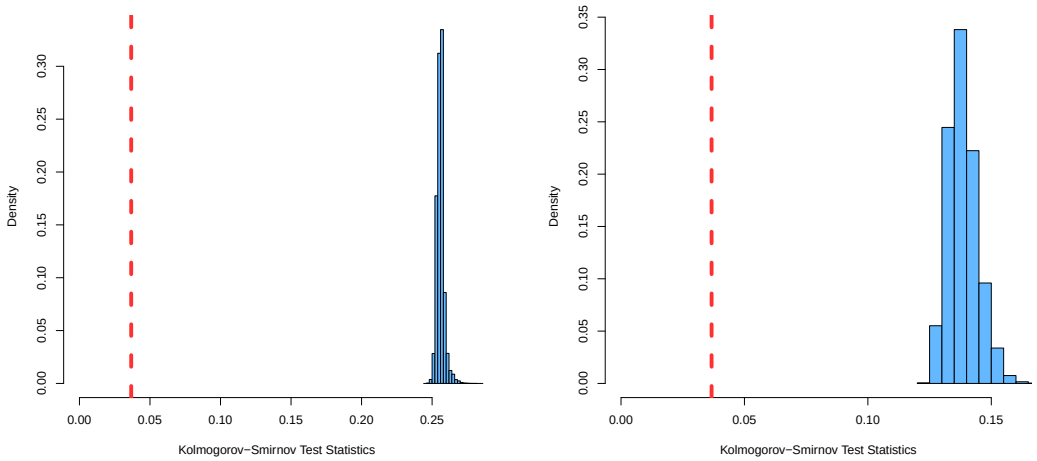
Figure C.10: Percentage change in List 1 responses

Note: Histogram of the percentage change between the second part and the first part of the experiment, defined as the difference between the utility of an allocation in List 1 Part 2 and the utility of the same allocation in List 1 Part 1, divided by the utility in List 1 Part 1. This graph includes the 21 allocations in List 1 for all 106 participants, for a total of 2226 observations.

Appendix D: Monte Carlo Simulation

The purpose of the Monte Carlo simulations is to compare the responses provided by our participants to the *inequity aversion list* with randomly generated values. We test the null hypothesis that the distribution of responses provided by participants in our sample and the distribution of random responses are equal. This test is replicated one million times, using a Kolmogorov-Smirnov test, with a different set of randomly generated data each time. Every iteration compares List 1 (Part 1) with the simulated random data.

Figure D.1 summarizes the results from the million tests for two types of random simulated data. On the left, figure D.1a shows a distribution of the test statistics generated in each iteration of the test when the random values are drawn from an uniform distribution over the support [1,7], corresponding to the range of values in which utility is expressed. On the right, figure D.1b shows the distribution of test statistics when the random values are generated from a normal distribution with the same mean and standard deviation as the distribution of responses provided by participants in our sample. In both cases, the dashed red line indicates the Kolmogorov-Smirnov critical value for $\alpha = 0.05$. Regardless of the underlying distribution of the random generated data, the null hypothesis is always rejected. Therefore, we conclude that the responses provided by participants in our sample to List 1 (Part 1) are statistically significantly different from random answers.



(a) r.v. $\sim U(1, 7)$

(b) r.v. $\sim N(\mu = 4.86, \sigma = 1.96)$

Figure D.1: Distribution for 1 million tests

Note: Red line indicates Kolmogorov-Smirnov critical value for $\alpha = 0.05$

Appendix E: Instructions

General instructions lists 1, 2 and 3

The following three tasks consist of lists of different allocations of money between you and the person with which you are paired. In each task you will be paired with a different person.

In these tasks you are asked to indicate your level of satisfaction with different allocations of money, on a scale of 1 to 7, where 1 corresponds to 'extremely dissatisfied,' and 7 corresponds to 'extremely satisfied.' Move the slider to the position that best represents your level of satisfaction with the corresponding allocation. The satisfaction value that corresponds to the slider's position can be viewed to the right of the scale, in order to assist you in rating your satisfaction accurately.

There are no wrong answers. Please answer carefully and do your best to indicate your exact level of satisfaction. You can move all the sliders as many times as you want before clicking the green button at the end of the page to proceed.

Payment:

At the end of the experiment, if one of the lists in the following three tasks is randomly selected for payment, one of the allocations on the list will be randomly chosen to determine the payments for you and the person with which you are paired. For example: if the first allocation on List 1 ("you get $20 and the other person gets $10") is randomly chosen to determine the payments in your pair, then either you or the person you are paired with will get $20, and the other person will get $10. You are equally likely to get $20 or $10.

It warrants mention that the position of the sliders in the following three tasks have no implication whatsoever on the final payment, should one of these tasks be chosen for final payment.

Ultimatum instructions

Your answers in the following two tasks may affect the final payment, should one of these tasks be chosen for final payment.

TASK: Triangle

You will make two decisions in this task. In the first decision, you are endowed with $20, and you decide how much to send to another participant (you have the option to send any whole dollar amount, $0 and $20 included.) That participant can either accept or reject your offer. If they accept, they keep the sent amount, and you keep the remaining endowment ($20 minus the amount sent.) However, if they reject your offer, you both get $0.

In the second decision, the roles are inverted and you decide the minimum amount you are willing to accept from another participant.

Payment:

At the end of the experiment, if this task is chosen for payment, one of the decisions you made will be randomly chosen to be implemented along with that of another participant. So, for instance, if your first decision was to send an amount $X, and that is larger than the minimum amount the other participant decided they were willing to accept, they get $X and you get $20 - $X. Otherwise, you both get zero.

Note that your second decision can also be randomly chosen for payment. In that case, the decision chosen for the other participant will be the first one.

You will go over two examples before making any decision.

Example 1:

Suppose your first decision is to send $2 to the other participant, and your second decision is to accept at least $1. Suppose the other participant makes the exact same decisions.

If this task is chosen for payment and your first decision is randomly assigned to be implemented for the payment:

How much money do you receive for this task at the end of the experiment?
(Do not include the $5 show up fee)

How much money does the other participant receive for this task at the end of the experiment?
(Do not include the $5 show up fee)

Example 2:

Suppose your first decision is to send $18 to the other participant, and your second decision is to accept at least $15. Suppose the other participant sends $10 as their first decision, and accepts at least $13 as their second decision.

If this task is chosen for payment and your second decision is randomly assigned to be implemented for the payment:

How much money do you receive for this task at the end of the experiment?
(Do not include the $5 show up fee)

How much money does the other participant receive for this task at the end of the experiment?
(Do not include the $5 show up fee)

>>

Sender's decision in Ultimatum

Now that you have practiced with the examples, please submit below your answers for the Triangle task. For your reference, you will find a copy of the task instructions at the bottom.

First decision:

How much do you want to send to another participant?

0

1

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

18

19

20

Receiver's decision in Ultimatum

Second decision:

What is the minimum amount you are willing to accept?

- 0
- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10
- 11
- 12
- 13
- 14
- 15
- 16
- 17
- 18
- 19
- 20

Dictator instructions

TASK: Diamond

This task has two players, one active and the other passive; player A makes decisions while player B does not. You will make decisions as player A, and later, a random draw will determine whether you are assigned the outcome of player A or player B.

Player A decides between either keeping $20 for him/herself while leaving player B with $0 (zero), or an equal payment for both players that varies from $0-$0 to $20-$20 as indicated in the table below, under the second column: 'Option B.'

As player A, you must first decide between keeping the $20 for yourself or splitting the payment. If you decide to split the payment, you must indicate the minimum amount you are willing to accept as an equal payment instead of keeping $20 for yourself. By indicating the minimum equal amount you are willing to accept, you are indicating you prefer option B for that amount and all higher equal split amounts, and you prefer option A for all lower equal split amounts.

For example, using the table below, if someone chooses option A for the entire table, they are not willing to split the payment. If someone chooses option A for rows 0-5 and then option B for rows 6-20, they are indicating that he or she is not willing to split up until a $6 payment for each player, at which point they are willing to split through row 20.

Row	Option A	Option B
0	Player A: $20, Player B: $0	Player A: $0, Player B:$0
1	Player A: $20, Player B: $0	Player A: $1, Player B:$1
2	Player A: $20, Player B: $0	Player A: $2, Player B:$2
3	Player A: $20, Player B: $0	Player A: $3, Player B:$3
4	Player A: $20, Player B: $0	Player A: $4, Player B:$4
5	Player A: $20, Player B: $0	Player A: $5, Player B:$5
6	Player A: $20, Player B: $0	Player A: $6, Player B:$6
7	Player A: $20, Player B: $0	Player A: $7, Player B:$7
8	Player A: $20, Player B: $0	Player A: $8, Player B:$8
9	Player A: $20, Player B: $0	Player A: $9, Player B:$9
10	Player A: $20, Player B: $0	Player A: $10, Player B:$10
11	Player A: $20, Player B: $0	Player A: $11, Player B:$11
12	Player A: $20, Player B: $0	Player A: $12, Player B:$12
13	Player A: $20, Player B: $0	Player A: $13, Player B:$13
14	Player A: $20, Player B: $0	Player A: $14, Player B:$14
15	Player A: $20, Player B: $0	Player A: $15, Player B:$15
16	Player A: $20, Player B: $0	Player A: $16, Player B:$16
17	Player A: $20, Player B: $0	Player A: $17, Player B:$17
18	Player A: $20, Player B: $0	Player A: $18, Player B:$18
19	Player A: $20, Player B: $0	Player A: $19, Player B:$19
20	Player A: $20, Player B: $0	Player A: $20, Player B:$20

Payment:

At the end of the experiment, if this task is chosen for payment, you and the person you are paired with will be randomly assigned to the roles of player A and player B. One person in the pair will be A, the other person will be B. Then ONE row will be randomly selected and player A's choice for that row will be implemented.

You will go over two examples before making any decision.

Dictator quiz

Example 1:

Suppose as player A you choose row 3 as the minimum amount you are willing to accept of an equal payment instead of keeping $20 for yourself, and the person you are paired with chooses row 19.

Suppose this task is chosen for payment. Also, suppose you are randomly assigned the role of player A for the payment and the randomly chosen row is 3:

How much money do you receive for this task at the end of the experiment? (Do not include the $5 show up fee)

How much money does the person you are paired with receive for this task at the end of the experiment? (Do not include the $5 show up fee)

Example 2:

Suppose as player A you choose row 15 as the minimum amount you are willing to accept of an equal payment instead of keeping $20 for yourself, and the person you are paired with chooses row 5.

Suppose this task is chosen for payment. Also, suppose you are randomly assigned the role of player B for the payment and the randomly chosen row is 4:

How much money do you receive for this task at the end of the experiment? (Do not include the $5 show up fee)

How much money does the person you are paired with receive for this task at the end of the experiment? (Do not include the $5 show up fee)

>>

Dictator's decision

Now that you have practiced with the examples, please submit below your answers for the Diamond task. For your reference, you will find a copy of the task instructions at the bottom.

As player A, the minimum amount you choose as equal split instead of $20 for you and $0 for player B is:

In other words, you choose Option B over Option A starting in row:

- 0
- 1
- 2
- 3
- 4
- 5
- 6
- 7
- 8
- 9
- 10
- 11
- 12
- 13
- 14
- 15
- 16
- 17
- 18
- 19
- 20
- I choose option A for the entire table

References

Alesina, A., Di Tella, R., & MacCulloch, R. (2004). Inequality and happiness: are europeans and americans different? *Journal of public economics*, *88*(9-10), 2009–2042.

Andreoni, J., & Miller, J. (2002). Giving according to garp: An experimental test of the consistency of preferences for altruism. *Econometrica*, *70*(2), 737–753.

Ashley, R., Ball, S., & Eckel, C. (2010). Motives for giving: A reanalysis of two classic public goods experiments. *Southern Economic Journal*, *77*(1), 15–26.

Bellemare, C., Kröger, S., & van Soest, A. (2011). Preferences, intentions, and expectation violations: A large-scale experiment with a representative subject pool. *Journal of Economic Behavior & Organization*, *78*(3), 349–365.

Bellemare, C., Kröger, S., & Van Soest, A. (2008). Measuring Inequity Aversion in a Heterogeneous Population Using Experimental Decisions and Subjective Probabilities. *Econometrica*, *76*(4), 815-839.

Benjamin, D. J., Cooper, K. B., Heffetz, O., & Kimball, M. S. (2019). Self-reported wellbeing indicators are a valuable complement to traditional economic indicators but aren't yet ready to compete with them. *Behavioural Public Policy, Forthcoming*.

Benjamin, D. J., Heffetz, O., Kimball, M. S., & Rees-Jones, A. (2012). What do you think would make you happier? what do you think you would choose? *American Economic Review*, *102*(5), 2083–2110.

Benjamin, D. J., Heffetz, O., Kimball, M. S., & Rees-Jones, A. (2014). Can marginal rates of substitution be inferred from happiness data? evidence from residency choices. *American Economic Review*, *104*(11), 3498–3528.

Benjamin, D. J., Heffetz, O., Kimball, M. S., & Szembrot, N. (2014). Beyond happiness

and satisfaction: Toward well-being indices based on stated preference. *American Economic Review*, *104*(9), 2698–2735.

Bertrand, M., & Mullainathan, S. (2001). Do people mean what they say? implications for subjective survey data. *American Economic Review*, *91*(2), 67–72.

Blake, P., McAuliffe, K., Corbit, J., Callaghan, T., Barry, O., Bowie, A., ... others (2015). The ontogeny of fairness in seven societies. *Nature*, *528*(7581), 258.

Blanchflower, D. G., & Oswald, A. J. (2004). Well-being over time in britain and the usa. *Journal of public economics*, *88*(7-8), 1359–1386.

Blanco, M., Engelmann, D., & Normann, H. T. (2011). A within-subject analysis of other-regarding preferences. *Games and Economic Behavior*, *72*(2), 321–338.

Blount, S. (1995). When social outcomes aren't fair: The effect of causal attributions on preferences. *Organizational behavior and human decision processes*, *63*(2), 131–144.

Bolton, G. E., & Ockenfels, A. (2000). Erc: A theory of equity, reciprocity, and competition. *American economic review*, *90*(1), 166–193.

Bolton, G. E., & Zwick, R. (1995). Anonymity versus punishment in ultimatum bargaining. *Games and Economic behavior*, *10*(1), 95–121.

Bosman, R., & Van Winden, F. (2002). Emotional hazard in a power-to-take experiment. *The Economic Journal*, *112*(476), 147–169.

Brañas-Garza, P., Galizzi, M. M., & Nieboer, J. (2018). Experimental and self-reported measures of risk taking and digit ratio (2d: 4d): evidence from a large, systematic study. *International Economic Review*, *59*(3), 1131–1157.

Brandts, J., & Solà, C. (2001). Reference points and negative reciprocity in simple sequential games. *Games and Economic Behavior*, *36*(2), 138–157.

Breza, E., Kaur, S., & Shamdasani, Y. (2018). The morale effects of pay inequality. *The Quarterly Journal of Economics*, *133*(2), 611–663.

Brodeur, A., & Flèche, S. (2019). Neighbors' income, public goods, and well-being. *Review of Income and Wealth*, *65*(2), 217–238.

Butler, J. V. (2016). Inequality and relative ability beliefs. *The Economic Journal*, *126*(593), 907–948.

Card, D., Mas, A., Moretti, E., & Saez, E. (2012). Inequality at work: The effect of peer salaries on job satisfaction. *American Economic Review*, *102*(6), 2981–3003.

Charness, G., & Grosskopf, B. (2001). Relative payoffs and happiness: an experimental study. *Journal of Economic Behavior & Organization*, *45*(3), 301–328.

Charness, G., & Rabin, M. (2000). Social preferences: Some simple tests and a new model. *Working paper, University of California*.

Charness, G., & Rabin, M. (2002). Understanding social preferences with simple tests. *The Quarterly Journal of Economics*, *117*(3), 817–869.

Chatterjee, P., Rose, R. L., & Sinha, J. (2013). Why money meanings matter in decisions to donate time and money. *Marketing Letters*, *24*(2), 109–118.

Cherry, T. L., Frykblom, P., & Shogren, J. F. (2002). Hardnose the dictator. *American Economic Review*, *92*(4), 1218–1221.

Clark, A. E., Frijters, P., & Shields, M. A. (2008). Relative income, happiness, and utility: An explanation for the easterlin paradox and other puzzles. *Journal of Economic literature*, *46*(1), 95–144.

Clark, A. E., Kristensen, N., & Westergård-Nielsen, N. (2009). Job satisfaction and co-worker wages: Status or signal? *The Economic Journal*, *119*(536), 430–447.

Clark, A. E., & Senik, C. (2011). Is happiness different from flourishing? cross-country evidence from the ess. *Revue d'économie politique*, *121*(1), 17–34.

Clark, A. E., Westergård-Nielsen, N., & Kristensen, N. (2009). Economic satisfaction and income rank in small neighbourhoods. *Journal of the European Economic Association*, *7*(2-3), 519–527.

Cohn, A., Fehr, E., Herrmann, B., & Schneider, F. (2014). Social comparison and effort provision: Evidence from a field experiment. *Journal of the European Economic Association*, *12*(4), 877–898.

Cullen, Z., & Perez-Truglia, R. (2018). How much does your boss make? the effects of salary comparisons. *NBER Working Paper Series, No. 24841, August 2018*.

D'Ambrosio, C., & Frick, J. R. (2012). Individual wellbeing in a dynamic perspective. *Economica*, *79*(314), 284–302.

Daruvala, D. (2010). Would the right social preference model please stand up! *Journal of Economic Behavior & Organization*, *73*(2), 199–208.

Deaton, A., & Stone, A. A. (2013). Two happiness puzzles. *American Economic Review*, *103*(3), 591–97.

Deci, E. L., & Ryan, R. M. (2000). The" what" and" why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological inquiry*, *11*(4), 227–268.

Dolan, P., Layard, R., & Metcalfe, R. (2011). Measuring subjective well-being for public policy.

Dube, A., Giuliano, L., & Leonard, J. (2019). Fairness and frictions: The impact of unequal raises on quit behavior. *American Economic Review*, *109*(2), 620–63.

Duckworth, A. L., & Yeager, D. S. (2015). Measurement matters: Assessing personal qualities other than cognitive ability for educational purposes. *Educational Researcher*, *44*(4), 237–251.

Duesenberry, J. S. (1949). Income, saving, and the theory of consumer behavior.

Dufwenberg, M., & Kirchsteiger, G. (2004). A theory of sequential reciprocity. *Games and economic behavior*, *47*(2), 268–298.

Easterlin, R. A. (1974). Does economic growth improve the human lot? some empirical evidence. In *Nations and households in economic growth* (pp. 89–125). Elsevier.

Easterlin, R. A. (2013). Happiness, growth, and public policy. *Economic Inquiry*, *51*(1), 1–15.

Easterlin, R. A., McVey, L. A., Switek, M., Sawangfa, O., & Zweig, J. S. (2010). The happiness–income paradox revisited. *Proceedings of the National Academy of Sciences*, *107*(52), 22463–22468.

Eckel, C., & Gintis, H. (2010). Blaming the messenger: Notes on the current state of experimental economics. *Journal of Economic Behavior & Organization*, *73*(1), 109–119.

Ellingsen, T., & Johannesson, M. (2004). Promises, threats and fairness. *The Economic Journal*, *114*(495), 397–420.

Engelmann, D., & Strobel, M. (2004). Inequality aversion, efficiency, and maximin preferences in simple distribution experiments. *American economic review*, *94*(4), 857–869.

Falk, A., Fehr, E., & Fischbacher, U. (2003). On the nature of fair behavior. *Economic inquiry*, *41*(1), 20–26.

Falk, A., & Fischbacher, U. (2006). A theory of reciprocity. *Games and economic behavior*, *54*(2), 293–315.

Fehr, E., Bernhard, H., & Rockenbach, B. (2008). Egalitarianism in young children. *Nature*, *454*(7208), 1079.

Fehr, E., & Fischbacher, U. (2003). The nature of human altruism. *Nature*, *425*(6960), 785.

Fehr, E., Naef, M., & Schmidt, K. M. (2006). Inequality aversion, efficiency, and maximin preferences in simple distribution experiments: Comment. *American Economic Review*, *96*(5), 1912–1917.

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, *114*(3), 817–868.

Fehr, E., & Schmidt, K. M. (2006). The economics of fairness, reciprocity and altruism–experimental evidence and new theories. *Handbook of the economics of giving, altruism and reciprocity*, *1*, 615–691.

Fershtman, C., Gneezy, U., & List, J. A. (2012). Equity aversion: Social norms and the desire to be ahead. *American Economic Journal: Microeconomics*, *4*(4), 131–44.

FitzRoy, F. R., Nolan, M. A., Steinhardt, M. F., & Ulph, D. (2014). Testing the tunnel effect: comparison, age and happiness in uk and german panels. *IZA Journal of*

European Labor Studies, *3*(1), 24.

Frank, R. H. (2016). *Success and luck: Good fortune and the myth of meritocracy.* Princeton University Press.

Frey, B. S., & Stutzer, A. (2002). What can economists learn from happiness research? *Journal of Economic literature*, *40*(2), 402–435.

Gagné, M. (2003). The role of autonomy support and autonomy orientation in prosocial behavior engagement. *Motivation and emotion*, *27*(3), 199–223.

Gale, J., Binmore, K. G., & Samuelson, L. (1995). Learning to be imperfect: The ultimatum game. *Games and economic behavior*, *8*(1), 56–90.

Galeotti, F., Montero, M., & Poulsen, A. (2018, 08). Efficiency versus Equality in Bargaining. *Journal of the European Economic Association*. Retrieved from `https://doi.org/10.1093/jeea/jvy030` (jvy030) doi: 10.1093/jeea/jvy030

Goeree, J. K., & Holt, C. A. (2000). Asymmetric inequality aversion and noisy behavior in alternating-offer bargaining games. *European Economic Review*, *44*(4-6), 1079–1089.

Green, F. (2010). Well-being, job satisfaction and labour mobility. *Labour Economics*, *17*(6), 897–903.

Grimm, V., & Mengel, F. (2011). Let me sleep on it: Delay reduces rejection rates in ultimatum games. *Economics Letters*, *111*(2), 113–115.

Güth, W., Huck, S., & Müller, W. (2001). The relevance of equal splits in ultimatum games. *Games and Economic Behavior*, *37*(1), 161–169.

Güth, W., Kliemt, H., & Ockenfels, A. (2003). Fairness versus efficiency: an experimental study of (mutual) gift giving. *Journal of Economic Behavior & Organization*, *50*(4), 465–475.

Hara, K., Adams, A., Milland, K., Savage, S., Callison-Burch, C., & Bigham, J. P. (2018). A data-driven analysis of workers' earnings on amazon mechanical turk. In *Proceedings of the 2018 chi conference on human factors in computing systems* (pp. 1–14).

Hauser, D. J., & Schwarz, N. (2016). Attentive turkers: Mturk participants perform

better on online attention checks than do subject pool participants. *Behavior research methods*, *48*(1), 400–407.

He, H., & Wu, K. (2016). Choice set, relative income, and inequity aversion: an experimental investigation. *Journal of Economic Psychology*, *54*, 177–193.

Hedegaard, M., Kerschbamer, R., Müller, D., & Tyran, J.-R. (2019). Distributional preferences explain individual behavior across games and time. *Available at SSRN 3388681*.

Helliwell, J., Layard, R., & Sachs, J. (2012). World happiness report.

Hirschman, A. O., & Rothschild, M. (1973). The changing tolerance for income inequality in the course of economic development: With a mathematical appendix. *The Quarterly Journal of Economics*, *87*(4), 544–566.

Hoffman, E., McCabe, K., & Smith, V. L. (1996). Social distance and other-regarding behavior in dictator games. *The American economic review*, *86*(3), 653–660.

Houser, D., & Xiao, E. (2010). Inequality-seeking punishment. *Economics Letters*, *109*(1), 20–23.

Ifcher, J., Zarghamee, H., & Graham, C. (2018). Local neighbors as positives, regional neighbors as negatives: Competing channels in the relationship between others' income, health, and happiness. *Journal of health economics*, *57*, 263–276.

Johansson-Stenman, O., Carlsson, F., & Daruvala, D. (2002). Measuring future grandparents' preferences for equality and relative standing. *The Economic Journal*, *112*(479), 362–383.

Kahneman, D. (2011). *Thinking, fast and slow.* Macmillan.

Kahneman, D., & Deaton, A. (2010). High income improves evaluation of life but not emotional well-being. *Proceedings of the national academy of sciences*, *107*(38), 16489–16493.

Kamas, L., & Preston, A. (2015). Can social preferences explain gender differences in economic behavior? *Journal of Economic Behavior & Organization*, *116*, 525–539.

Kerschbamer, R. (2015). The geometry of distributional preferences and a non-parametric identification approach: The equality equivalence test. *European economic review*,

76, 85–103.

Kingdon, G. G., & Knight, J. (2007). Community, comparisons and subjective well-being in a divided society. *Journal of Economic Behavior & Organization*, *64*(1), 69–90.

Konow, J., & Earley, J. (2008). The hedonistic paradox: is homo economicus happier? *Journal of public Economics*, *92*(1-2), 1–33.

Krueger, A. B., & Schkade, D. A. (2008). The reliability of subjective well-being measures. *Journal of public economics*, *92*(8-9), 1833–1845.

Kuhn, P., Kooreman, P., Soetevent, A., & Kapteyn, A. (2011). The effects of lottery prizes on winners and their neighbors: Evidence from the dutch postcode lottery. *American Economic Review*, *101*(5), 2226–47.

Kuziemko, I., Buell, R. W., Reich, T., & Norton, M. I. (2014). "last-place aversion": Evidence and redistributive implications. *The Quarterly Journal of Economics*, *129*(1), 105–149.

Kuziemko, I., Norton, M. I., Saez, E., & Stantcheva, S. (2015). How elastic are preferences for redistribution? evidence from randomized survey experiments. *American Economic Review*, *105*(4), 1478–1508.

Levine, D. K. (1998). Modeling altruism and spitefulness in experiments. *Review of economic dynamics*, *1*(3), 593–622.

List, J. A., Shaikh, A. M., & Xu, Y. (2016). Multiple hypothesis testing in experimental economics. *Experimental Economics*, 1–21.

Loewenstein, G. F., Thompson, L., & Bazerman, M. H. (1989). Social utility and decision making in interpersonal contexts. *Journal of Personality and Social psychology*, *57*(3), 426.

Ludwig, J., Duncan, G. J., Gennetian, L. A., Katz, L. F., Kessler, R. C., Kling, J. R., & Sanbonmatsu, L. (2013). Long-term neighborhood effects on low-income families: Evidence from moving to opportunity. *American economic review*, *103*(3), 226–31.

Luttmer, E. F. (2005). Neighbors as negatives: Relative earnings and well-being. *The*

Quarterly journal of economics, *120*(3), 963–1002.

McBride, M. (2010). Money, happiness, and aspirations: An experimental study. *Journal of economic behavior & organization*, *74*(3), 262–276.

Mogilner, C. (2010). The pursuit of happiness: Time, money, and social connection. *Psychological Science*, *21*(9), 1348–1354.

Murphy, R. O., Ackermann, K. A., & Handgraaf, M. (2011). Measuring social value orientation. *Judgment and Decision making*, *6*(8), 771–781.

Nielsen, K., & Rehbeck, J. (2019). Are axioms normative? eliciting axiom preferences and resolving conflicts with lottery choices. *Working Paper*.

Peterson, A. R., & Matthew, S. (1983). *A mood scale for survey research. american marketing association's educators' conference.* Chicago: AMA.

Pfeffer, J., & DeVoe, S. E. (2009). Economic evaluation: The effect of money and economics on attitudes about volunteering. *Journal of Economic Psychology*, *30*(3), 500–508.

Pigou, A. (1920). The economics of welfare. london: Macmillan and co., ltd.

Quoidbach, J., Dunn, E. W., Petrides, K. V., & Mikolajczak, M. (2010). Money giveth, money taketh away: The dual effect of wealth on happiness. *Psychological science*, *21*(6), 759–763.

Rabin, M. (1993). Incorporating fairness into game theory and economics. *The American economic review*, 1281–1302.

Rees, J. C. (1956). Mill and his early critics.

Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003). The neural basis of economic decision-making in the ultimatum game. *Science*, *300*(5626), 1755–1758.

Segal, U., & Sobel, J. (2007). Tit for tat: Foundations of preferences for reciprocity in strategic settings. *Journal of Economic Theory*, *136*(1), 197–216.

Segal, U., & Sobel, J. (2008). A characterization of intrinsic reciprocity. *International Journal of Game Theory*, *36*(3-4), 571–585.

Senik, C. (2004). When information dominates comparison: Learning from russian subjective panel data. *Journal of Public Economics*, *88*(9-10), 2099–2123.

Senik, C. (2008). Ambition and jealousy: Income interactions in the 'old'europe versus the 'new'europe and the united states. *Economica*, *75*(299), 495–513.

Smith, R. H., Diener, E., & Wedell, D. H. (1989). Intrapersonal and social comparison determinants of happiness: A range-frequency analysis. *Journal of personality and social psychology*, *56*(3), 317.

Snowberg, E., & Yariv, L. (2018). *Testing the waters: Behavior across participant pools* (Tech. Rep.). National Bureau of Economic Research.

Solnick, S. J., & Hemenway, D. (1998). Is more always better?: A survey on positional concerns. *Journal of Economic Behavior & Organization*, *37*(3), 373–383.

Stevenson, B., & Wolfers, J. (2008). *Economic growth and subjective well-being: Reassessing the easterlin paradox* (Tech. Rep.). National Bureau of Economic Research.

Stiglitz, J. E., Sen, A., Fitoussi, J.-P., et al. (2009). *Report by the commission on the measurement of economic performance and social progress.* Citeseer.

Stone, A. A., & Mackie, C. E. (2013). *Subjective well-being: Measuring happiness, suffering, and other dimensions of experience.* National Academies Press.

Sutter, M., Zoller, C., & Glätzle-Rützler, D. (2019). Economic behavior of children and adolescents–a first survey of experimental economics results. *European Economic Review*, *111*, 98–121.

Tasch, W., & Houser, D. (2018). *Social preferences and social curiosity.* (CESifo Working Paper Series No. 7132. Available at SSRN: https://ssrn.com/abstract=3235299)

Van Praag, B. M., & Ferrer-i-Carbonell, A. (2004). *Happiness quantified: A satisfaction calculus approach.* Oxford University Press.

Vohs, K. D., Mead, N. L., & Goode, M. R. (2006). The psychological consequences of money. *Science*, *314*(5802), 1154–1156.

Xiao, E., & Houser, D. (2005). Emotion expression in human punishment behavior. *Proceedings of the National Academy of Sciences*, *102*(20), 7398–7401.

Yang, Y., Onderstal, S., & Schram, A. (2016). Inequity aversion revisited. *Journal of Economic Psychology*, *54*, 1–16.

Zaleskiewicz, T., Gasiorowska, A., Kesebir, P., Luszczynska, A., & Pyszczynski, T. (2013). Money and the fear of death: The symbolic power of money as an existential anxiety buffer. *Journal of Economic Psychology*, *36*, 55–67.

Zarghamee, H., Ifcher, J., Houser, D., & Diaz, L. (2020). The relative income effect: An experiment. *Experimental Economics*. Retrieved from `https://doi.org/10.1007/s10683-020-09648-w`

Zeckhauser, R. (1991). *Strategy and choice*. Mit Press Cambridge.

Zhou, X., Vohs, K. D., & Baumeister, R. F. (2009). The symbolic power of money: Reminders of money alter social distress and physical pain. *Psychological Science*, *20*(6), 700–706.

Curriculum Vitae

Lina Díaz was born in Colombia, where she received her Bachelor (2010) and Master (2012) degrees, both in Economics from Universidad de Los Andes. Before coming to the United States, she worked as a research assistant at Universidad del Rosario from 2011 to 2013.