The University of Nottingham

School of Economics

Essays on Inequality Aversion, Conditional Cooperation, and Punishment

Benjamin Beranek, M.Sc.

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy

October 2016

Abstract

Recent developments in behavioral and experimental economics have shown that many people display other-regarding motives, in addition to self-interest. These social preferences – which include inequality aversion, conditional cooperation, and motives for punishment – make sense of a number of phenomena left unexplained by standard economic theory. This thesis is a collection of studies examining social preferences using the tools of experimental economics. Chapter 1 introduces the thesis explaining our goals and methods, as well as previewing our substantive contributions. Chapter 2 reports an experiment designed both to replicate and extend previous studies. We elicit and compare *stated* and *revealed* inequality aversion at the individual-level for subjects drawn from three different subject pools. Chapter 3 investigates whether inequality aversion, as modelled by Fehr and Schmidt (1999), explains free riding and conditional cooperation in a public good game. Chapter 4 investigates whether observed variation in the directionality of punishment between two subject pools can be explained by the "Culture of Honor" hypothesis. Chapter 5 concludes.

Acknowledgements

Completing this thesis would not have been possible without the support of a great many people. First and foremost, I am grateful to my supervisors Professor Robin Cubitt and Professor Simon Gächter. I have not met anyone with as much vision or boundless enthusiasm as Professor Simon Gächter. I appreciate his intellectual curiosity, his willingness to take risks on new ideas, and his ability to capture the essence of complexity. Professor Robin Cubitt is the most precise and thoughtful economist I know. I am inspired by his careful attention to detail (often shown through the precise script of his handwritten notes along the margins of a manuscript) and the deep respect he holds for the ideas of others. Their strengths as supervisors complement one another and I have benefited tremendously learning from them both individually and collectively.

I am also grateful to Dr. Lucas Molleman. He has been a great source of support and encouragement throughout my PhD, but especially in relation to Chapter 4. His technical prowess has enabled, his thoughtful comments have improved, and his eye for figures has helped better illustrate this work.

Many thanks also to Professor Dirk Engelmann and Dr. Simone Quercia for their practical help in providing zTree files, instructions, and datasets as well as their extremely useful comments. I am also grateful to Professor Ayla Oğuş Binaltı for her warm hospitality and efforts in facilitating my research at Izmir University of Economics.

A great many people contributed to this work in various ways. Thanks to Professor George Loewenstein, Professor Urs Fischbacher, and Professor Ernst Fehr for their helpful comments during their visits to Nottingham. Many members of the Centre for Decision Research and Experimental Economics (CeDEx) have offered valuable support including Professor Chris Starmer, Professor Abigail Barr, Dr. Daniele Nosenzo, and Dr. Elke Renner. I'm grateful to Professor Sourafel Girma and Dr. Markus Eberhardt for their econometric guidance. I thank Ms. Sue Berry, Ms. Sarah Nolan and Ms. Suzanne Robey for their practical support over the years. I am also very grateful to the Professor Kevin Lee and the School of Economics for their support during my studies.

The studies in this thesis were funded by the European Research Council Project Cooperation (ERC-AdG 295707 COOPERATION) and I am grateful for their support. Even more valuable than the financial support was the vibrant research community which this ERC funding enabled. Our weekly ERC Discussion Group meetings over lunch were a place where many of the ideas in this thesis were developed. Many thanks to Simon for including me in that group and to the members of that group for the countless conversations and their great insights into these areas. Thanks to Dr. Jonathan Schulz, Dr. Ori Weisel, Dr. Antonio Alonso Arechar, Dr. Roberto Hernan Gonzalez, Dr. Felix Kölle, Dr. Ozan Işler, and Dr. Marcus Giamattei as well as Simon and Lucas.

The School of Economics at the University of Nottingham has been a wonderful place to complete a PhD and do this research. In particular, I am grateful to Mr. Pawel Smietanka, Mr. Geoffrey Castillo, and Ms. Ashira Perera with whom I have had the privilege of working alongside for the duration of my PhD. Your friendship and support means so much to me. Many thanks also to my colleagues Mr. Basile Boulay, Ms. Martina Magli, Ms. Marta Ronchetti, Mr. Tim DeStefano, Mr. Donneil Cain, Ms. Emily Whitehouse, Ms. Hande Erkut, Ms. Georgina Michailidou, Ms. Erika Barbosa, Ms. Karen Roopnarine, Mr. Till Weber, Mr. Ruslan Shichman, Mr. Tom Lane, Mr. Andreas Friedl, Dr. Kian Ong, Dr. Jörg Weber, Dr. Samantha Torrance, Dr. Emilija Timmis, Dr. Jonathan Timmis, Dr. Zahra Murad, and Dr. Francesco Fallucchi.

Beyond the School of Economics, I am grateful to so many friends in Nottingham for their support including Dr. Adam Algar, Ms. Leslie Bode, Mr. Darren Macrow, Ms. Louise Phillips, Dr. Iker Perez Lopez, Ms. Jyothika Kumar, Professor Jan Bradley, Mr. Rowan Thomas, Mr. David Keasey, Mr. Harry Villa, Mr. Andrew Wilkinson, Ms. Rosa Lavelle-Hill, Dr. Barnabas King, Dr. Rosie Adsley, Dr. Rebecca Haines, Dr. Jillian Wall, Mr. Jon Tan, Mr. Henry Curran, and Mr. Tom Devas.

Finally, I must acknowledge my family for their unfailing love and support. Many thanks to my parents, Terry and Marcia; my sister, Katie; my brother-in-law, Patrick; my nephew, Charley; and my niece, Gretel. I am so grateful for you all.

Contents

Abstract 3

Acknowledgements 4

List of Figures 10

List of Tables 12

Chapter 1. Introduction 14

1.1 Motivation 14

1.2 Our Contribution 17

1.3 Methods 19

Chapter 2 Stated and revealed inequality aversion in three subject pools 21

2.1 Introduction 21

2.2 Methods 25

2.3 Results 30

2.3.1 Stated Inequality Aversion 30

2.3.2 Revealed Inequality Aversion 32

2.3.3 Relationship between Stated and Revealed Preferences 35

2.3.4 The role of socio-demographics and guilt proneness for inequality aversion 37

2.4 Discussion and Conclusion 39

2.5 Supplementary Materials 45

Chapter 3 Does inequality aversion explain free riding and conditional cooperation? 80

3.1 Introduction 80

3.2 Research Strategy 83

3.2.1 Preferences 83

3.2.2 Environment 85

3.2.3 Predictions 86

3.3 Experimental Design 88

3.3.1 Parameter of Advantageous Inequality Aversion, βi 90

3.3.2 Conditional Cooperation and Free Rider Strategy Types 92

3.3.3 One-Shot Direct Response Public Goods Game with Belief Elicitation 94

3.4 Results 96

3.4.1 Preliminary Findings 97

3.4.2 Does Inequality Aversion explain Free Riding and Conditional Cooperation? 100

3.4.3 Further Analyses and Other Explanations 106

3.5 Conclusions 112

3.6 Supplementary Materials 116

Chapter 4 Culture of Honor Revisited: Examining Punishment Behavior of US Southerners in Social Dilemmas 145

4.1 Introduction 145

4.2 Culture of Honor 148

4.3 Study 1 152

4.3.1 Method 152

4.3.2 Results and Discussion 156

4.4 Study 2 163

4.4.1 Methods 163

4.4.2 Results and Discussion 166

4.5 General Discussion 169

4.6 Supplementary Materials 174

Chapter 5 Conclusions 182

Bibliography 187

List of Figures

Figure 2.1. Stated Preferences on Aggregate and Individual Levels.....31

Figure 2.2. Joint α and β Distributions 35

Figure 2.3. Joint a and α Distributions.....36

Figure 2.4. Joint b and β Distributions.....37

Figure 2.5. Example Screen shots of zTree Scenario Decision Screens Sequence Two 62

Figure 2.6. The original quadratic LBT89 functional form expressed as a social utility curve emphasizing the importance of relative payoffs.....66

Figure 2.7. Linear Social Utility Curves.....69

Figure 2.8. Expanded Versions of Figure 2.1B to Figure 2.1D in the Main Text 71

Figure 2.9. Expanded Versions of Figure 2.2 in the Main Text 77

Figure 3.1. Distribution of the parameters of advantageous inequality aversion (βi).....98

Figure 3.2. FGF Type Distribution.....99

Figure 3.3. Histograms of β Values by FGF Types 103

Figure 3.4. Histogram of Difference between Unconditional and Conditional Contributions 108

Figure 3.5. Contributions and βi Values 133

Figure 3.6 Contributions and βi Values with Mean Values Split into Quintiles 136

Figure 3.7. Contribution Schedules per Subject 138

Figure 4.1. Game tree of Study 1 155

Figure 4.2. First mover punishment rates for Southerners and Non-Southerners 157

Figure 4.3. Game tree of Study 2 165

Figure 4.4. Punishment rates when retaliation is impossible 167

Figure 4.5. First Mover punishment rates when retaliation is possible for Southerners and Non-Southerners. 178

Figure 4.6. Punishment rates when retaliation is impossible for randomly selected Southern and Non-Southern punishers. 181

List of Tables

Table 2.1. Distribution of α and β values.....33

Table 2.2. OLS regression analysis of demographic and psychological determinants of stated and revealed parameters of inequality aversion. ..38

Table 2.3. Subject Pool Details.....47

Table 2.4. Summary Scenario Sequences and Resulting OLS Estimates. 56

Table 2.5. Summary statistics and correlations between Loewenstein et al. Dispute Conditions 68

Table 2.6. Comparing the constructed parameters to the estimated parameters. 70

Table 2.7. Kruskal-Wallis Tests comparing parameters of stated inequality aversion by populations. 73

Table 2.8. Wilcoxon-Mann-Whitney test comparing parameters of stated inequaliity aversion between populations. 73

Table 2.9. Group Chi^2 Tests Comparing α and β Categories.....74

Table 2.10. Kruskal-Wallis Tests comparing parameters of revealed inequality aversion by populations. 75

Table 2.11. Wilcoxon-Mann-Whitney test comparing parameters of revealed inequality aversion between populations. 75

Table 2.12. Incidences of Multiple Switching in Izmir. 76

Table 3.1. Association between Types 101

Table 3.2. Probit Analysis of Perfect Conditional Cooperators 111

Table 3.3. Standard Subject Pool Statistics 131

Table 3.4. Subject Pool Psychological Measures 132

Table 3.5. *p*-values of pair wise Wilcoxon rank sum tests (two-tailed) comparing contributions between quintiles 137

Table 3.6. Probit Analysis of Perfect Conditional Cooperators, Coefficients from All Models 140

Table 3.7. Likelihood-Ratio Tests comparing Models. 141

Table 3.8. Probit Analysis of Perfect Conditional Cooperators, Marginal Effects for Model 7 142

Table 4.1. Second Mover Punishment Behavior in Study 1. 159

Table 4.2. Social demographic characteristics as reported in Study 1 ... 177

Table 4.3. Tukey-Kramer Pairwise Pairwise Comparisons of the Different Categories of Second Mover Punishment for Southerners 179

Table 4.4. Tukey-Kramer Pairwise Pairwise Comparisons of the Different Categories of Second Mover Punishment for Non-Southerners 180

Table 4.5. Social demographic characteristics as reported in Study 2 ... 180

Chapter 1.

Introduction

1.1 Motivation

Many of the core postulates in the discipline of economics have been challenged over the last few decades. Increasing globalization after the fall of communism, incredible new developments in technology, and unanticipated events like the rise of terrorism and the Great Recession have changed the ways we think about how our economies interrelate and function on national and international levels. On the opposite end of the macro-micro spectrum, recent experimental economics research has challenged the accounts of individual time, risk, and social preferences put forward by standard economic theory and has led to the development of alternative behavioral economic theory which better explains the observed data. This thesis is a collection of three papers about individual social preferences in which we use the tools of experimental economics to test elements of this new behavioral economic theory. In particular, we are interested in the concepts of inequality aversion, conditional cooperation, and punishment.

Both Chapter 2 and Chapter 3 draw extensively on the Fehr and Schmidt (1999) model of inequity aversion. According to this theory, an individual's utility is made up of several components. Like standard economic theory, an individual derives positive utility from her own material payoffs. Unlike standard economic theory, individuals can derive disutility from inequality in payoffs between themselves and their comparators. As Fehr and Schmidt (and subsequently many others) show, the addition of the possibility of disutility from inequality in payoffs is a possible explanation for a large number of phenomena unexplainable by standard economic theory.

In Chapter 2, we elicit parameters of inequality aversion using two different techniques. We replicate portions of the Loewenstein, Thompson, & Bazerman (1989) study that Fehr and Schmidt (1999) cite as an inspiration for their model of inequity aversion. In their non-incentivized study, Loewenstein, Thompson, & Bazerman (1989) had their subjects read vignettes and indicate their satisfaction for a variety of outcomes in order to generate social utility curves. We did likewise and refer to the parameters of inequality aversion elicited in this manner as measures of *stated* inequality aversion. We also had the same subjects take part in incentivized economic games following the methodology introduced by Blanco, Engelmann, & Normann (2011) to elicit what we call measures of *revealed* inequality aversion. This within-subject design enables us to compare stated and revealed inequality aversion at the individual level, as well as to investigate the Fehr and Schmidt (1999) model of inequity aversion and its motivating assumptions.

In Chapter 3, we explore the concept of conditional cooperation beside inequality aversion. In a social dilemma, an individual must choose between

costly cooperation and free riding. The socially optimal outcome is achieved when everyone cooperates, but an individual achieves the maximal material payoff by free riding. Choosing to cooperate leads to a social return that is greater than one's own private cost; however, the social dilemma arises because the fraction of the social return a cooperator receives is smaller than her private cost of contribution. The prediction of standard economic theory is therefore under-provision of public goods (Samuelson, 1954), but a whole host of evidence shows that people sometimes cooperate nonetheless (Ledyard, 1995; Chaudhuri, 2011). Using an innovative experimental design, Fischbacher, Gächter, & Fehr (2001) observed that a large proportion of people choose to conditionally cooperate in social dilemmas matching (or at least responding positively to) the contributions of others.[1] In Chapter 3, we explore whether inequality aversion is the driving force behind conditional cooperation. More specifically, we see whether predictions derived from the Fehr and Schmidt (1999) model of inequity aversion explain observed free riding and conditional cooperation in a public good game. The idea behind this hypothesis is that if person *i* contributes less than person *j*, it puts *i* at a material advantage relative to *j*. The extent of *i*'s aversion to the resulting inequality determines their willingness to do this.

In Chapter 4, we then shift our focus towards punishment. Fehr & Gächter (2000) show how the introduction of costly punishment can dramatically increase the levels of cooperation in a repeated public good game.[2] Punishment

[1] Studies which follow on from this result reporting conditional cooperation in social dilemmas include [e.g., Henrich et al. (2005), Herrmann et al. (2008) and Blanco et al. (2011)].
[2] Chaudhuri (2011) provide a helpful survey of the subsequent literature.

seems to be a mechanism whereby the problem of free riding can be overcome and its introduction leads to the realization of outcomes that are closer to the socially optimal. However, Herrmann, Thoni, & Gächter (2008) have shown that not all cultures respond in the same way to the introduction of punishment. In some societies punishment is a prosocial tool with strong directionality towards free riders and its introduction leads to enhanced overall cooperation. In others societies, punishment is used with less directionality and it is sometimes even used to punish those who contribute the most.

In Chapter 4, we examine one possible explanation for why that might be the case. The *Culture of Honor hypothesis* states that individuals from particular cultures respond aggressively to any challenge to or impugning of their honor. This response ensures that one does not gain a reputation as someone easily taken advantage of. In the context of a social dilemma with punishment and counter-punishment possibilities, it could be the case that those who receive punishment interpret it as a challenge to their honor and therefore the culture of honor requires them to respond in kind. We devise a novel experiment to test whether or not this culture of honor hypothesis can explain the observed variation in the directionality of punishment in one particular society in which the Culture of Honor has been shown to operate: the United States.

1.2 Our Contribution

Overall, we add new insights into social preferences. In Chapter 2, we replicate the stated inequality aversion findings of Loewenstein et al. (1989) and find these results to be rather robust across disparate subjects pools. While these findings which motivated the development of the Fehr and Schmidt (1999)

model of inequity aversion are quite robust, we find that revealed inequality aversion measures are much less so. Using the Blanco et al. (2011) methodology to elicit revealed inequality aversion, we find widespread violations of the assumption that advantageous inequality is preferred to disadvantageous inequality. Further, we find weak correlation between the stated and revealed measures of inequality aversion suggesting that the scenario tasks and the laboratory games are perhaps measuring different things.

In regards to whether inequality aversion explains free riding and conditional cooperation, we find in Chapter 3 that the Fehr and Schmidt (1999) model of inequity aversion has only limited predictive success. Those subjects identified as perfect conditional cooperators tend to have strong (revealed) aversion to inequality, but those identified as free riders often do not have weak inequality aversion. The Fehr and Schmidt (1999) model also does not account for the majority of subjects who are classified as imperfect conditional cooperators or other types.

Finally, in Chapter 4's examination of the impact of the culture of honor in explaining variation in punishment behavior in the United States of America, we find – contrary to our predictions – that US Southerners (who belong to a Culture of Honor society) punish defectors less frequently than their Non-Southern counterparts as first punishers in a social dilemma with sequential punishment. In light of this unexpected result, we conjectured that it was the possibility of retaliation that lowered the rate at which US Southerners were punishing. Removing the possibility of retaliation in our second study removed the difference in punishment rates between Southerners and Non-Southerners.

1.3 Methods

Each of the main chapters in this thesis makes use of the empirical tools of experimental economics. Across these three chapters, subjects in our experiments participate in varying combinations of different economic games such as the Blanco, Engelmann, & Normann (2011) modified dictator and ultimatum games, the Fischbacher, Gächter, & Fehr (2001) public good game played with the strategy method, the Fischbacher, Gächter, & Quercia (2012) public good game with belief elicitation, and our novel simultaneous social dilemmas with punishment. In all of these economic games, subjects are paid according to the decisions that they and their coplayers made. These games are presented to subjects abstractly without any context.

Our analysis also makes use of non-incentivized measures. As mentioned above, subjects in Chapter 2 read vignettes and indicate their satisfaction of a variety of outcomes. In both Chapter 2 and Chapter 3, we make use of various psychological measures including Gosling et al. (2003)'s Ten-Item Personality Inventory (TIPI) measuring the Big Five personality domains, Cohen et al. (2011)'s Guilt and Shame Proneness Scale, Perugini et al. (2003)'s Personal Norm of Reciprocity Questionnaire, and Frederick (2005)'s Cognitive Reflection Test. These non-incentivized measures complement the findings from our economic games and provide further validation of the reported results.

Additionally, we recognize that one limitation of experimental economics research is that those who participate in it are often student subjects from European or North American universities. A growing number of studies have found that these subjects have unique characteristics which differentiates them

from others (Henrich, Heine, & Norenzayan, 2010; Barr et al., 2009; Herrmann, Thoni, & Gächter, 2008; Gächter & Schulz, 2016). Seeking to address this concern towards the generalizability of experimental economics findings, we include in our studies subjects from a culturally different society (in Chapter 2) and beyond the university (in Chapter 2 and Chapter 4).

More specifically, we collect the data for Chapter 3 and part of Chapter 2 using lab experiments conducted at the University of Nottingham as part of a larger research project. For greater generalizability of our results, we collect additional data for Chapter 2 using lab experiments at Izmir University of Economics in Izmir, Turkey and online from subjects across the United States of America using the Amazon MTurk platform. For Chapter 3, our analysis includes data we reported in Chapter 2 – specifically, the revealed preference data of advantageous inequality aversion as well as a psychological measure (GASP) from the lab experiments conducted at the University of Nottingham – together with data from additional economic games and psychological measures we did not report in Chapter 2. Finally for Chapter 4, we use data collected from subjects across the United States of America using the Amazon MTurk platform.

Chapter 2

Stated and revealed inequality aversion in three subject pools[3]

2.1 Introduction

Inequality aversion, the dislike of unequal outcomes, has become established as one of the core postulates of behavioural economics. Although discussion of equity concerns is by no means new (e.g., Adams 1965; Selten 1978), the recent literature took off with publication of formal models of inequality averse preferences by Bolton (1991), Bolton & Ockenfels (2000) and Fehr and Schmidt (1999), with the latter paper providing the most widely applied model. In this paper, we are concerned with the range of empirically relevant parameters of inequality aversion. We concentrate on aversion to inequality in bilateral monetary comparisons, as in Fehr and Schmidt's theory.

[3] This chapter draws on joint work with Robin Cubitt and Simon Gächter published in the Journal of the Economic Science Association (2015).

To our knowledge, Loewenstein, Thompson, & Bazerman, (1989) is the first paper to provide systematic evidence on this. They presented their subjects with (hypothetical) life-like scenarios that involved distributions of money between the subject and a comparator and asked the subjects to rate their satisfaction with those distributions. We refer to these measurements as *stated preferences* because they do not involve choices, but un-incentivized assessments of satisfaction. On the basis of these ratings, Loewenstein, et al. estimated 'social utility functions' and found that subjects dislike inequality when it is to their advantage and when it is to their disadvantage. However, aversion to disadvantageous inequality was considerably stronger than aversion to advantageous inequality. Fehr and Schmidt use this observation (pp. 821, 823-4) to justify their distinctive assumption that disadvantageous inequality aversion (measured in their model by a parameter called α) is at least as strong as advantageous inequality aversion (measured by a parameter called β). A key subsequent step taken by Blanco, Engelmann, & Normann (2011) was to provide *individual-level* measures of α and β, respectively, by using subjects' choices in two particular games. Thus, importantly, their measures of the parameters reflect *revealed preferences*, in the traditional sense.

In this paper, we replicate the Loewenstein, et al. experiments using updated versions of their scenarios to elicit social utility functions; and we replicate Blanco, et al.'s measurement of α and β using their games and procedures.[4] In the latter case, like Blanco, et al., we will observe the joint

[4] We are aware of three further studies, by Dannenberg, Riechmann, Sturm, & Vogt (2007), Dannenberg, Riechmann, Sturm, & Vogt (2012), and Teyssier (2012), that elicit revealed

distribution of α and β and so be able to reassess the extent to which elicited values are consistent with Fehr and Schmidt's assumption that $\alpha \geq \beta$. However, our most novel contribution is that we link stated and revealed inequality aversion at the individual level: for each subject, our experimental design yields parameters of stated disadvantageous and advantageous inequality aversion obtained with methods akin to those of Loewenstein, et al., and values of α and β revealed by choices using Blanco, et al.'s methods. As they refer to the same inequalities but are obtained with different methods, we use *a* and *b* to denote the stated preference analogues of α and β, respectively.

If inequality aversion is a general sentiment triggered across different situations, then stated and revealed measures should be positively correlated across individuals. If they are, measurements of stated and revealed preferences cross-validate each other. If they are not, this would call into question how strongly the findings of Loewenstein, et al. could support the modelling assumptions of Fehr and Schmidt. Further, this finding would complement the findings of studies which report correlation across games between different measures of revealed pro-sociality such as Yamagishi et al. (2013), Dariel & Nikiforakis (2014), and Peysakhovich, Nowak, & Rand (2014).

We also investigate the association between inequality aversion and proneness to guilt, which we measure using the GASP (guilt and shame proneness) scale of Cohen et al. (2011). This is particularly relevant to aversion

---

preferences using methods akin to Blanco, et al. but with research questions distinct from ours. We are not aware of previous replications of Loewenstein, et al.'s seminal work.

to advantageous inequality, the parameter of which (here b or β) is often referred to as the "guilt" parameter (e.g., Blanco, et al., p. 322).

Our data are from three subject pools, two of them drawn from the student bodies of the University of Nottingham (UK) and Izmir University of Economics (Turkey) and the third from the American online workforce of Amazon Mechanical Turk (MTurk).[5] Across all subject pools, 717 people participated in our experiments. Apart from differences pertaining to the subject pools, the experimental procedures were essentially uniform.

Our main results are as follows. Notwithstanding some differences in intensity, stated inequality averse preferences are qualitatively similar in all three subject pools in that, like Loewenstein, et al., we find that $a \geq b$ and $b \geq 0$ for most subjects. This provides strong support for the findings of Loewenstein, et al. that inspired Fehr and Schmidt's theory. However, the support for some aspects of that theory itself is weaker, as we find violations at the individual and the median level in all subject pools of the assumption that $\alpha \geq \beta$. We find only weak positive correlation between a and α. Correlation between b and β is significantly positive and exceeds that for a and α in all subject pools. We find females are more averse than males to advantageous inequality and that there is an association between inequality aversion and the GASP measure of proneness to guilt and shame. Although there are some differences between our findings

[5] MTurk is an online labor market. Horton, Rand, & Zeckhauser (2011) discusses its usefulness for experimental economics. A more sceptical view is expressed by Chandler, Mueller, & Paolacci (2014).

from different subject pools, they are mostly not important for central tendencies of parameters, especially once other factors are controlled for.

2.2 Methods

For brevity, we focus in this section on the main features of our experimental designs, relegating technical and procedural details, instructions and scenario texts to the Supplementary Materials, Sections 2.5.1.1, 2.5.1.2, and 2.5.1.3. Each subject completed all of the game tasks described in this section, a selection of scenario tasks, the GASP task and some other tasks with no feedback until all tasks had been completed. We describe the games here in terms of "points", as we did to subjects. At the end of the experiment, points from one game were converted to cash.

The core of this study is the two-person version of the Fehr and Schmidt (1999) model of inequality aversion:

$$U_i = x_i - \alpha_i \max\{x_j - x_i, 0\} - \beta_i \max\{x_i - x_j, 0\}, \qquad 2.1$$

$$j \neq i, \alpha_i \geq \beta_i; 1 > \beta_i \geq 0.$$

In this functional form, U_i denotes person *i*'s utility, x_i person *i*'s monetary payoff and x_j the other person's monetary payoff. The parameter α_i governs *i*'s disutility from disadvantageous inequality, i.e. from x_i falling short of x_j; and the parameter β_i governs *i*'s disutility from advantageous inequality, i.e. from x_i exceeding x_j. A core assumption of Fehr and Schmidt is that advantageous inequality has less of a negative impact on overall utility than disadvantageous inequality of the same magnitude, i.e. $\alpha_i \geq \beta_i$. A central goal of our paper is to provide fresh estimates of the joint distribution of α_i and β_i.

A subject's parameter β_i of advantageous inequality aversion is elicited using a *Modified Dictator game* that Blanco, et al. introduced and which we implement in the same way. The dictator has to make 21 decisions, each a choice between the distribution (20 points for self, 0 points for other) and an equal distribution (*x* points for self, *x* points for other), where "other" refers to a passive player. The equal distributions increased in increments of 1 point from (0, 0) to (20, 20) in the obvious notation. Subjects were told that, should this game be selected as payoff relevant, then they would be paired with a randomly selected player, allocated the role of dictator or passive recipient, and one of the dictator's 21 decisions would be realized. As explained by Blanco, et al. (p. 325-326), the dictator's β_i parameter is theoretically determined by the equal distribution $(\tilde{x}_i, \tilde{x}_i)$ which he regards as good as the distribution (20, 0). From equation (1), $U_i(20, 0) = U_i(\tilde{x}_i, \tilde{x}_i)$ if, and only if, $20 - 20\beta_i = \tilde{x}_i$. Thus,

$$\beta_i = 1 - \frac{\tilde{x}_i}{20}. \qquad 2.2$$

Following Blanco, et al., we predict that, as *x* rises in steps, subjects will switch (once) from choosing (20, 0) over (*x*, *x*) to making the opposite choice; and, given this, we approximate $\tilde{x}_i$ as the average of the highest *x* for which (20, 0) is chosen and the lowest *x* for which (*x*, *x*) is chosen. $\beta_i = 1$ (resp. 0) is assigned to a subject who always (resp. never) chooses the equal option.

Following Blanco, et al. (p. 325) (and in line with a suggestion of Fehr and Schmidt) behavior in the *Ultimatum game* of Güth, Schmittberger, & Schwarze (1982) can be used to elicit the parameter α_i of revealed disadvantageous inequality aversion. One player (the proposer) proposes to the other player (the responder) an allocation of a fixed sum (here of 20 points). Then, the responder

chooses between accepting the proposal and rejecting it. In the former case, the proposal is implemented; but, in the latter case, both participants receive 0 points. All subjects make decisions in both roles, using the strategy method for the responder's decision, so as to provide a response to all distributions that might be proposed. Subjects were told that, should this game be selected as payoff relevant, then they would be paired with a randomly selected player, assigned roles, and the payoffs would be realized should the responder have accepted the proposal according to the strategy method. If the responder had rejected the proposal then both would receive 0 points.

A subject's strategy in the role of responder yields an estimate of their α_i parameter. This is determined theoretically by the proposal $\tilde{s}_i$ at which the responder is just indifferent between accepting and rejecting. From Equation 2.1, $U_i(\tilde{s}_i, 20 - \tilde{s}_i) = \tilde{s}_i - \alpha_i(20 - \tilde{s}_i - \tilde{s}_i) = 0$ determines the point of indifference, and thus,

$$\alpha_i = \frac{\tilde{s}_i}{2(10-\tilde{s}_i)}. \qquad 2.3$$

Following Blanco, et al., when there is no more than one switch-point in the responder's strategy, we approximate $\tilde{s}_i$ with the average of the lowest accepted offer and the highest rejected offer. Subjects who do not reject any offers are assigned $\alpha_i = 0$ and those who reject all offers less favorable to them than an equal split are assigned $\alpha_i = 4.5$ (Blanco, et al., p. 325).

Our elicitation of α_i and β_i relies on subject i "switching" no more than once in the relevant sequence of choices for each parameter. Following Blanco, et al., any subject who switched multiple times in either of these sequences, and

is thus not *well behaved*, is excluded from the data reported in Section 2.3. We excluded 18% of subjects from Nottingham, 55% of subjects from Izmir, and 9% of subjects from MTurk. This compares to Blanco, et al. who excluded 15% of subjects for not being *well behaved*.

We turn now to the elicitation of *stated* preferences. In Studies 1 & 2 of Loewenstein et al. (1989), participants read various scenarios describing a range of possible distributions of outcomes, with the subject in the role of one of the affected parties in a bilateral dispute. Subjects ranked their satisfaction of outcomes on an 11-point scale. In our study, we follow the design of Loewenstein, et al., but use modernized scenario tasks. This part of the design is 2×2×2, varying the nature of the issue disputed (distributing the proceeds of an invention or of a plot of land between two parties), the prior relationship between the two parties (positive or negative), and whether it is gains or losses which are to be distributed. Thus, in total there are eight different scenarios. For each one, the task is to rate 21 distributions of payoffs for the subject and the other person described in the scenario. Each subject was presented with four different scenarios. For each subject, the resulting 84 ratings are used to estimate a "social utility" function of the same form as Equation 2.1 (plus the addition of a constant) to obtain estimates of their stated advantageous and disadvantageous inequality aversion parameters (called a_i and b_i). The estimation used OLS, with the subject's stated satisfaction as the dependent variable. By construction, this procedure produces a value of a_i and a value of b_i for each subject but, as explained above, subjects with non-well-behaved *revealed* preferences are excluded from Section 2.3. This guarantees that the

revealed and stated preferences reported are drawn from the same set of subjects.

Though we may expect positive rank correlation across individuals between a_i and α_i (resp. between b_i and β_i) if inequality aversion is a general sentiment across domains, the presence of such correlation is certainly not built in to the design. The scenario tasks and the games are quite different from one another: for example, only the latter are incentivised, the response modes are not the same, and the contexts described in the scenarios differ from those posed by the game instructions.

We are interested in the generalizability across subject pools of the findings on stated and revealed inequality aversion. Replication in a culturally different society and outside the university environment is important because there is mounting evidence that student subjects from European or North American universities often are quite special, when compared to others (Henrich, Heine, & Norenzayan, 2010; Barr et al., 2009; Herrmann, Thoni, & Gächter, 2008).

We conducted sessions at the University of Nottingham (n=104 students, all British); at Izmir University of Economics (n=206 students, all Turkish); and on the MTurk platform (n= 407 adult residents of the US). Culturally, there is significant distance between the UK and Turkey; and the MTurk sample differs from both university samples, especially by having greater variety of ages and education levels, and by being American.

In both university samples, the experiment was programmed in zTree (Fischbacher 2007) using, in the case of the games, zTree code and instructions provided by Blanco, et al.. In Nottingham, recruitment was done using ORSEE

(Greiner, 2015). In Izmir, recruiting required approaching students on campus. The MTurk experiments were conducted using online survey software Qualtrics on the MTurk platform. In all cases, the experiments were followed by post-experimental questionnaires, which elicited socio-demographic information, as well as the measure of guilt and shame proneness (Cohen et al., 2011).

2.3 Results

2.3.1 Stated Inequality Aversion

Our first result concerns the a_i and b_i parameters of stated inequality aversion derived from the scenarios. We constructed the social utility curves of Figure 2.1A by averaging the a_i and b_i parameters across each subject pool (and normalizing the constant to zero). Figure 2.1A shows averaged utility as a function of the difference between Own and Others payoff, when one's own payoff is zero. In all three of our subject pools, averaged utility is positively sloped in the region of disadvantageous inequality (to the left of zero on the horizontal axis) and negatively sloped in the region of advantageous inequality. Also, in all three of our subject pools and in line with Loewenstein, et al., the slope of averaged utility is greater in absolute value in the former region than in the latter, implying that disadvantageous inequality had a greater negative impact on satisfaction ratings than the corresponding advantageous inequality.

Figure 2.1A also shows that, especially in the region of advantageous inequality, the averaged social utility curves of the Nottingham and Izmir subject pools are quite similar to one another and to corresponding curves from the Loewenstein, et al. findings (see Supplementary Materials, Section 2.5.1.3.3). In contrast, the averaged social utility curves from the MTurk sample

show more pronounced aversion to both forms of inequality. Kruskal-Wallis tests confirm that there are statistically significant differences between our subject pools in b_i values ($\chi^2(2)=8.779$, p=0.0124) and, especially, in a_i values ($\chi^2(2)=23.858$, $p<0.001$), an issue to which we return in Section 2.4.

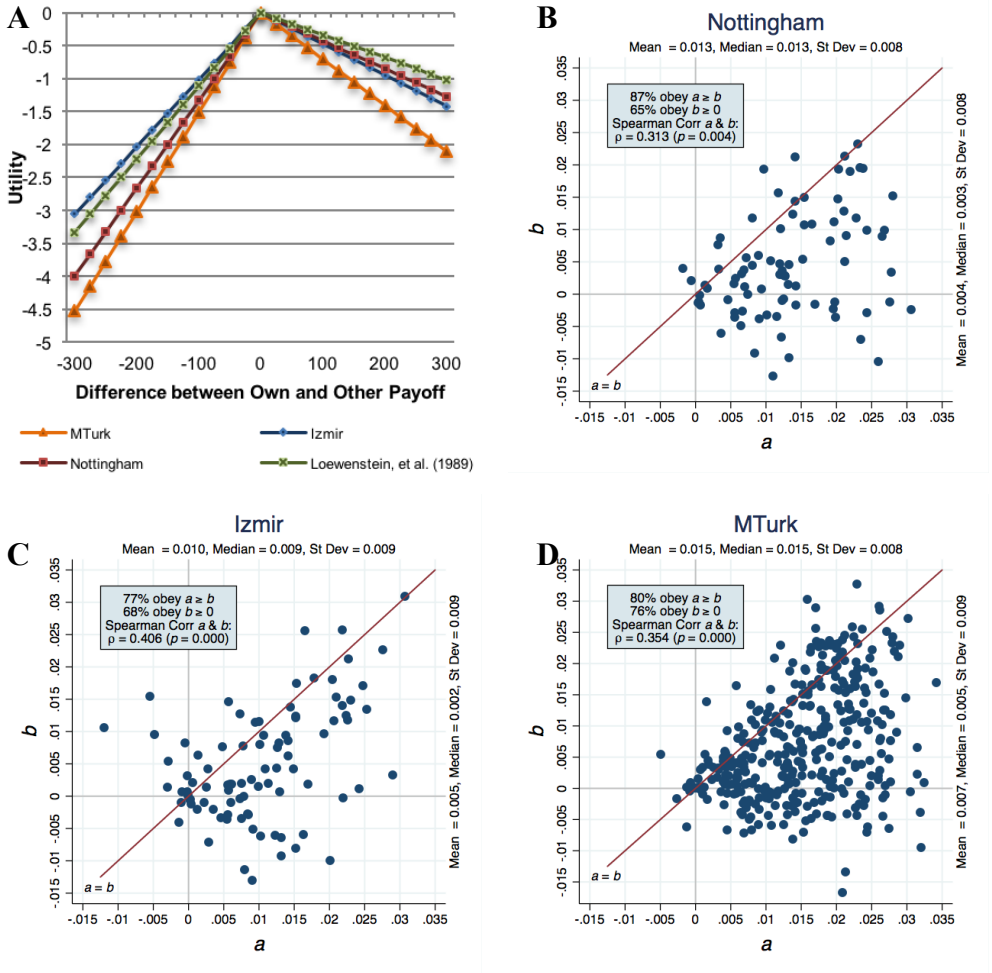


Figure 2.1. Stated Preferences on Aggregate and Individual Levels. Fig. 2.1A. Utility (satisfaction ratings) as a function of difference between own and other payoff in the scenario tasks. **Figure 2.1B to Figure 2.1D. Joint *a* and *b* distributions per subject pool.** Each dot represents a participant's a and b parameters as calculated from their stated preferences in the updated Loewenstein, et al (1989) scenario tasks. (The corresponding individual level data for Loewenstein, et. al. are not available.) Observations to the left of the $a = b$ line have $a < b$.

Figure 2.1B to Figure 2.1D illustrate the joint a_i and b_i distributions for each of our subject pools. Recall that a key assumption of Fehr and Schmidt is that $\alpha_i \geq \beta_i$ which they justify referring to the Loewenstein, et al results. In our notation, the corresponding finding to that of Loewenstein, et al. would be tendency for a_i to exceed b_i. We find strong support for this, as $a_i \geq b_i$ for 87%, 77%, and 80% of participants in the Nottingham, Izmir, and MTurk subject pools, respectively. However, a non-trivial minority (35%, 32% and 24% respectively) violate the condition $b_i \geq 0$, so displaying a stated preference *for* advantageous inequality. This finding is consistent with observation of the existence of spiteful preferences in related literature (e.g., Balafoutas, Kerschbamer, & Sutter, 2012; Iriberri & Rey-Biel, 2013).

Figure 2.1B to Figure 2.1D also report that a_i and b_i are positively correlated across individuals in each subject pool. In the pooled data, the corresponding Spearman rho is 0.3784 ($p<0.0001$).

2.3.2 Revealed Inequality Aversion

Table 2.1 shows the distribution of α_i (top panel) and β_i (lower panel) for each subject pool, using the categories of Blanco, et al.'s Table 2. We compare our observed distributions to Blanco, et al.'s interpretation (p. 326) of the distribution which Fehr and Schmidt deem plausible and to the distributions which Blanco, et al. themselves observe. Table 2.1 reports the relevant Chi^2-tests in each case, as well as the mean, median and standard deviation of each parameter in each subject pool.

The upper panel of Table 2.1 reveals that, in all three of our subject pools, values of α_i in the range of $\alpha_i < 0.4$ are substantially more frequent than in the

Blanco, et al. data (between 46% and 59% of our subjects have an $\alpha_i < 0.4$, compared to 31% in Blanco, et al.). Values of $\alpha_i \geq 4.5$ are, with the exception of Nottingham, also more frequent in our subject pools than in the Blanco, et al. data (7%, 24%, and 17%, in Nottingham, Izmir, and MTurk, respectively, compared to 13% in Blanco, et al.). Chi^2-tests confirm that all three of our subject pools differ significantly (at p=0.03 or lower) in respect of α_i from both the Fehr and Schmidt and the Blanco, et al. distributions of this parameter.

α	F&S	BEN	Nottingham	Izmir	MTurk
$\alpha < 0.4$	30%	31%	54%	59%	46%
$0.4 \leq \alpha < 0.92$	30%	33%	18%	12%	17%
$0.92 \leq \alpha < 4.5$	30%	23%	21%	5%	20%
$4.5 \leq \alpha$	10%	13%	7%	24%	17%
Mean		1.181	0.754	1.227	1.218
Median		0.611	0.269	0.026	0.410
Std. Dev.		1.488	1.198	1.884	1.670
Chi^2 test to F&S	χ^2 (3)	1.790	11.226	37.751	17.211
	p value	0.618	0.011	0.000	0.001
Chi^2 test to BEN	χ^2 (3)		9.014	24.933	9.699
	p value		0.029	0.000	0.021

β	F&S	BEN	Nottingham	Izmir	MTurk
$\beta < 0.235$	30%	29%	21%	16%	20%
$0.235 \leq \beta < 0.5$	30%	15%	25%	11%	19%
$0.5 \leq \beta$	40%	56%	54%	73%	61%
Mean		0.473	0.484	0.589	0.512
Median		0.525	0.525	0.575	0.525
Std. Dev.		0.310	0.290	0.315	0.302
Chi^2 test to F&S	χ^2 (2)	8.51	3.816	21.517	14.491
	p value	0.014	0.148	0.000	0.001
Chi^2 test to BEN	χ^2 (2)		2.729	5.033	3.109
	p value		0.256	0.081	0.211
Total Sample Size		72	104	206	407
Percent Well Behaved		85%	82%	45%	91%

Table 2.1. Distribution of α and β values. BEN refers to the Blanco, et al. (2011) observed distribution and F&S to Fehr and Schmidt. The data in these two columns and the row classifications are reproduced from Blanco, et al. (p. 325). Percent Well Behaved includes participants who had at most one switching point in the Ultimatum Game and at most one switching point in the Modified Dictator game. Only these participants are included in the analysis of this paper; all others are excluded.

In contrast, our distributions of β_i values appear similar to Blanco, et al.'s findings. Chi^2-tests find no significant difference from the distribution of β_i values reported by Blanco, et al. in our Nottingham and MTurk samples (p=0.26 and 0.21, respectively), and only a weakly significant difference in the Izmir sample (p=0.08). Comparing our distributions to those assumed by Fehr and Schmidt, using Chi^2-tests, reveals significantly different distributions in the Izmir and MTurk sample (p<0.01), but an insignificant difference between the Fehr-Schmidt distribution and that of our Nottingham sample (p=0.15).

Our next result concerns Fehr and Schmidt's assumption that $\alpha_i \geq \beta_i$. A first, aggregate level, take is provided by comparing the means and medians documented in Table 2.1. We find that the mean value of α_i is indeed larger than the mean β_i in all our subject pools (as in Blanco, et al.). However, the median α_i is *lower* than the median β_i in all our subject pools (unlike in Blanco, et al.).

Table 2.1 also shows notable variation in the percent of 'well-behaved' participants (as defined above) in each subject pool. In the Blanco, et al. subject pool, 85% of participants were well-behaved. Our Nottingham and MTurk subject pools displayed similar percentages of well-behaved participants (82% and 90% respectively), but only 45% of our Izmir sample met the criteria of well-behavedness.

The four panels of Figure 2.2 give the joint (α_i, β_i) distributions for the Blanco, et al. subject pool and for each of our subject pools. As was foreshadowed in the medians, we see many violations of the assumption that $\alpha_i \geq \beta_i$ in our subject pools. Whereas Blanco, et al. reported 38% of their participants violating this assumption, we find 55%, 59%, and 51% of

participants violating it in Nottingham, Izmir, and MTurk, respectively. Like Blanco, et al., we also find that α_i and β_i are uncorrelated in Nottingham and Izmir; in the MTurk sample the correlation between α_i and β_i is slightly (but significantly) negative. In the pooled data, the correlation is very slightly negative (rho = - 0.089; p=0.038).

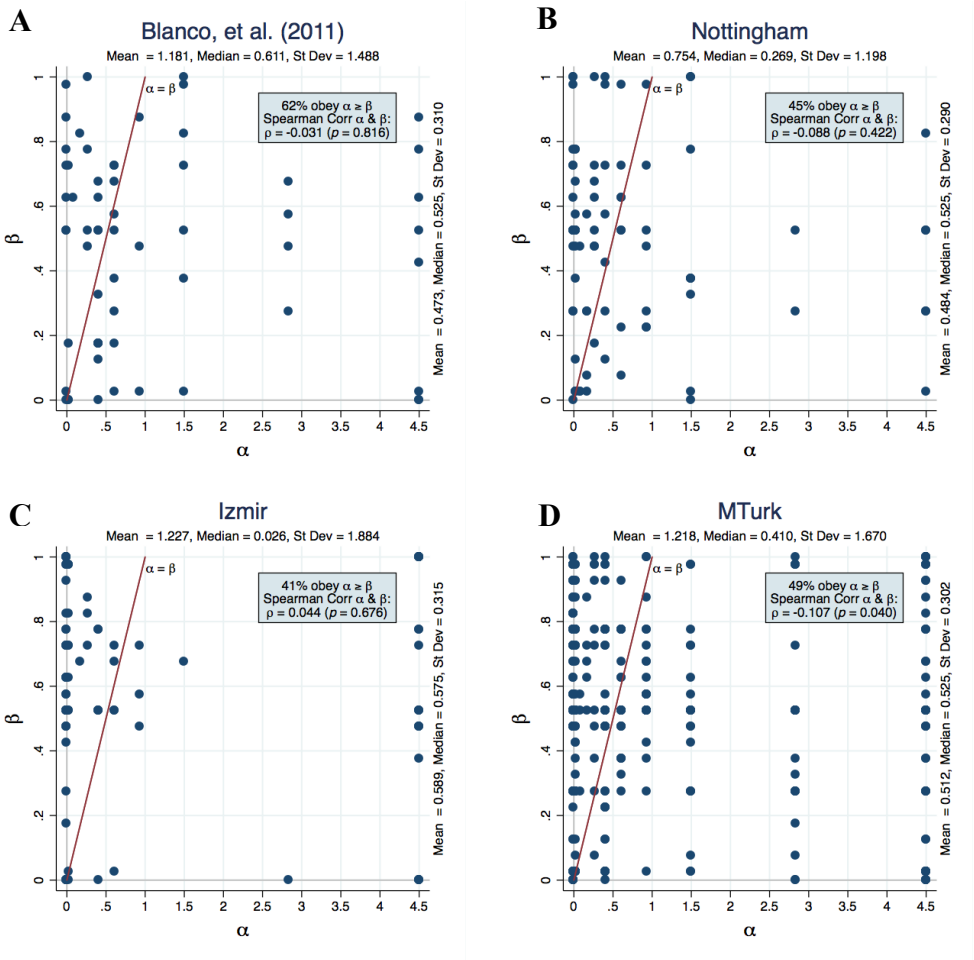


Figure 2.2. Joint α and β Distributions. Each dot represents a participant's α and β parameters as calculated from their revealed preferences in the Blanco, et al (2011) games. Observations to the left of the $\alpha = \beta$ line have $\alpha < \beta$ which violates the Fehr and Schmidt (1999) assumption.

2.3.3 Relationship between Stated and Revealed Preferences

Figure 2.3 shows the joint distribution of α_i (stated) and α_i (revealed) parameters of disadvantageous inequality aversion for each subject pool, with the associated Spearman's rho and its significance level. Surprisingly, there is

no significant correlation between a_i and α_i in the Izmir pool; and, though the correlation is statistically significant in the other two pools, it is only rather weakly positive, especially in the MTurk sample. In the pooled data, the correlation is slightly positive (rho = 0.132; p=0.002).

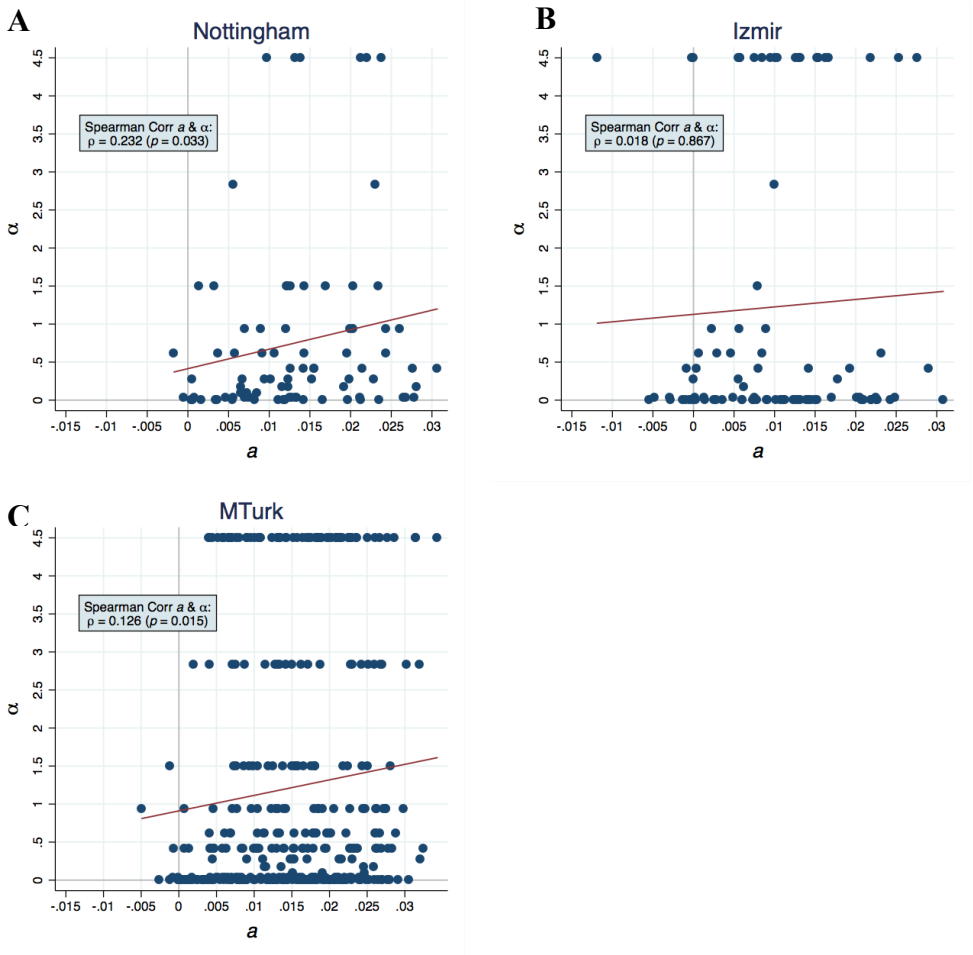


Figure 2.3. Joint a and α Distributions. Each dot represents a participant's a as calculated from their stated preferences and α parameters as calculated from their revealed preferences. The line results from the linear regression of α on a.

The corresponding materials for the joint distribution of b_i (stated) and β_i (revealed) parameters of advantageous inequality aversion are shown in Figure 2.4. For these parameters, the correlation is positive and statistically significant in all three subject pools. The degree of correlation is still quite modest, but

higher in each subject pool than for a_i and α_i. In the pooled data, the correlation is moderately positive (rho = 0.2785; p<0.001).

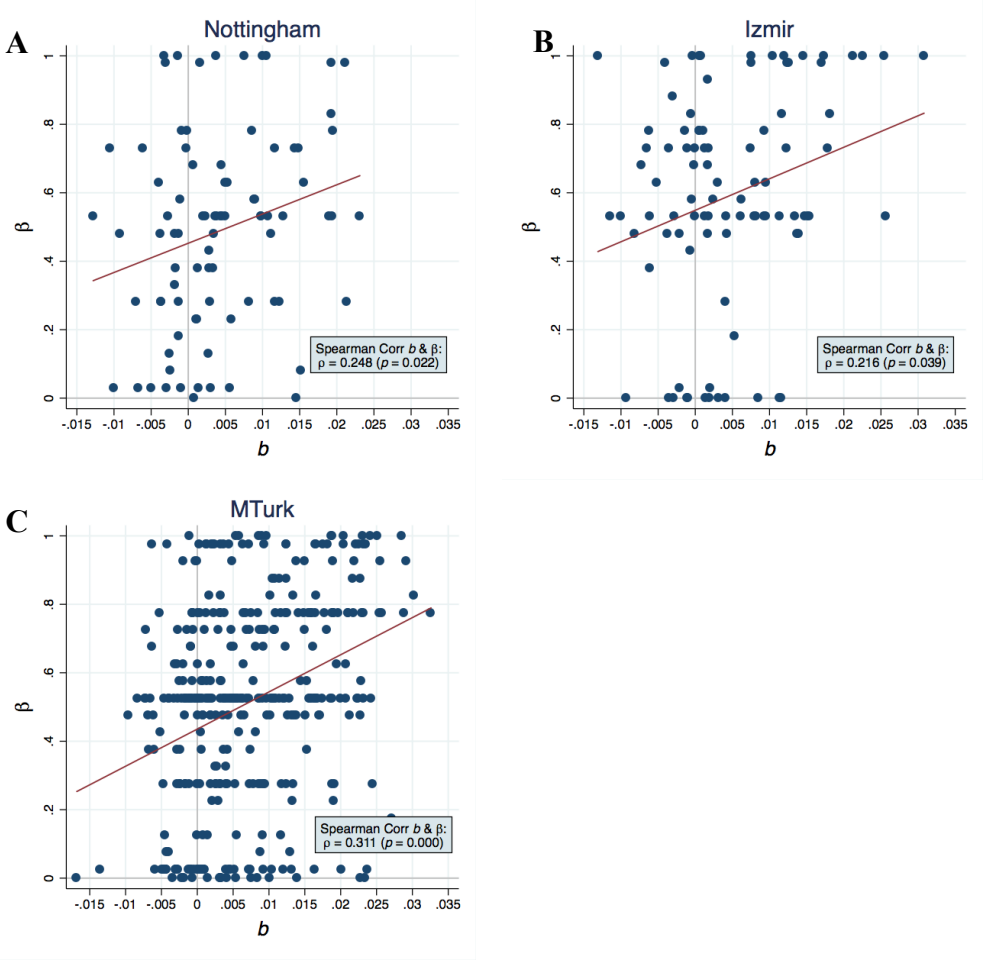


Figure 2.4. Joint *b* and *β* Distributions. Each dot represents a participant's *b* as calculated from their stated preferences and *β* parameters as calculated from their revealed preferences. The line results from the linear regression of *β* on *b*.

2.3.4 The role of socio-demographics and guilt proneness for inequality aversion

Finally, we pool the data from all three subject pools and separately regress our four measures of inequality aversion (a_i, b_i, α_i, β_i) on three standard socio-demographic variables – namely age, a female dummy, and a dummy for having some post-secondary education – and on a dummy for having studied Economics or Business, the GASP scale, and on dummies for Izmir and MTurk (the Nottingham subject pool being the omitted category). Across all our

subject pools there is considerable age variation (18-75 years), largely due to the MTurk population. Between 41% and 46% of participants were females, across the three subject pools. All our university student subjects and 82% of the MTurk workers have post-secondary education status, with 5% having studied Economics and 12% Business. Table 2.2 records the results of the regressions.

	a	b	α	β
Age	0.0002***	0.0002***	0.0130	0.0016
	(0.0000)	(0.0000)	(0.0086)	(0.0015)
Female	0.0014*	0.0045***	0.1006	0.0674***
	(0.0007)	(0.0007)	(0.1491)	(0.0267)
Post Secondary	-0.0020*	-0.0016	0.0670	0.0269
	(0.0011)	(0.0011)	(0.2308)	(0.0413)
Economics	-0.0021	-0.0017	-0.3219	-0.0939*
	(0.0015)	(0.0015)	(0.3098)	(0.0554)
Business	0.0002	-0.0019*	0.1591	-0.0505
	(0.0010)	(0.0010)	(0.2115)	(0.0378)
GASP	0.0013***	0.0016***	0.0329	0.0572***
	(0.0004)	(0.0004)	(0.0873)	(0.0156)
Izmir	-0.0041***	-0.0003	0.4332*	0.0753*
	(0.0012)	(0.0013)	(0.2540)	(0.0454)
MTurk	-0.0015	-0.0003	0.2382	-0.0239
	(0.0011)	(0.0012)	(0.2362)	(0.0423)
Constant	0.0060***	-0.0060***	0.2695	0.1571*
	(0.0023)	(0.0023)	(0.4690)	(0.0839)
N	546	546	546	546
F (8, 537)	8.88***	13.53***	1.46	5.33***
Adjusted R^2	0.1037	0.1553	0.0067	0.0597

Table 2.2. OLS regression analysis of demographic and psychological determinants of stated and revealed parameters of inequality aversion.

Age is a significant predictor of stated inequality aversion of both forms, but not of the corresponding revealed preference parameters. Compared with males, females state slightly stronger aversion to disadvantageous inequality aversion, but show significantly higher estimates for stated and revealed

advantageous inequality aversion. This result is consistent with the experimental evidence that females give more than men in dictator games (e.g., Eckel & Grossman, 1998; Engel, 2011). Participants who had studied either Economics or Business showed marginally significantly lower aversion to advantageous inequality (for Economics in revealed preference but for Business in stated preference). Finally, with the exception of α_i, all inequality aversion parameters are highly significantly positively correlated with GASP (higher scores indicate a greater proneness towards guilt and shame). The remaining subject pool difference that stands out once all these factors are controlled for is that Izmir subjects have significantly lower values of a_i.

2.4 Discussion and Conclusion

In terms of replication, our main results are as follows. The qualitative findings of Loewenstein, et al. appear rather robust in that the central tendencies of our *stated* preference data, in each subject pool, support the hypothesis of aversion to advantageous and disadvantageous inequality, with the latter the more keenly felt. Thus, these findings reinforce one of the main ingredients of Fehr and Schmidt's motivation for their model. In contrast, our findings on *revealed* preferences conform less closely to the assumptions of Fehr and Schmidt's model and to the findings of Blanco, et al., whose revealed preference techniques we use. We find widespread violation of Fehr and Schmidt's assumption that $\alpha_i \geq \beta_i$. Although our results on the distribution of the parameter β_i of aversion to advantageous inequality are similar to the corresponding findings of Blanco, et al., our distribution of the parameter α_i of disadvantageous inequality aversion differ markedly from that observed by Blanco, et al. Compared with them, we find a notably larger proportion of *low*

values of the parameter in all of our subject pools; and also a larger proportion of *high* values of the parameter, in particular in our Izmir sample.

Below, we comment further on two of our most striking findings - weak correlation between stated and revealed preferences and the frequent violation of Fehr and Schmidt's assumption that $\alpha_i \geq \beta_i$, just mentioned – and on differences between our subject pools.

We observe statistically significant positive rank correlations across individuals between (stated) b_i and (revealed) β_i parameters of advantageous inequality aversion in all three of our subject pools; and between (stated) a_i and (revealed) α_i parameters of disadvantageous inequality aversion in two of those pools. But, in all six cases, Spearman's rho was below 0.32, suggesting only a weak relationship. We can think of three possible reactions to this.

One possible perspective is that difference between stated and revealed preferences is an indication of "hypothetical bias" in the former, arising perhaps because subjects do not take un-incentivized tasks seriously or use them to indulge in cheap talk. A second possible perspective is that the difference between stated and revealed preferences, combined with greater conformity of the former to theoretical predictions, indicates that the scenario tasks "work" better, perhaps because subjects find them more recognizable or accessible than the stripped-down lab games. A third perspective is that the scenario tasks and the laboratory games both "work", but they measure slightly different things - in one case, an attitude and, in the other case, willingness to take a certain kind of action. These are correlated because there is an underlying propensity to act on ones attitudes. But, the correlation need not be strong, for example if the

propensity to act on ones attitudes is itself a trait whose strength varies across individuals.

To elaborate, the satisfaction ratings of the scenario tasks may indicate subjects' happiness with (or feelings about) different outcomes, whereas the Modified Dictator and Ultimatum games indicate subjects' willingness to sacrifice monetary payoffs in order to change the payoff of the other player in the game. This perspective chimes with the discussion of Blanco, et al. (Section 7) of their finding that the Fehr-Schmidt model, taken with parameter values elicited with their revealed preference methods, predicts the play of games other than those used in the elicitation less successfully at the individual level than at the aggregate level. They point out that willingness to give up money in order to change the other players' payoff may be sensitive to the nature of the game, as well as to the type of inequality faced.

These considerations are also relevant to our findings about the relative strength of aversion to advantageous and disadvantageous inequality. Even if adherence to some ethical codes might induce the opposite attitude, we would expect most subjects to be happier on receiving the larger part of some given unequal allocation between two people than on receiving the smaller part. If the satisfaction ratings of our scenario tasks are indicators of happiness, in this sense, then our stated preference findings strongly support this expectation. In contrast, our finding that a majority of subjects violate the assumption that $\alpha_i \geq \beta_i$ is a matter of revealed preferences. Viewed more narrowly, it is a matter of the trade-offs that subjects are willing to make in two particular games.

A subject assigned a low value of α_i is one who is reluctant to leave positive offers on the table when playing as respondent in the Ultimatum game. We report more instances of this than most previous studies, but reluctance to leave money on the table is not completely counter-intuitive behavior, even for a subject who feels unhappy about getting less than the proposer. And, of course, *homo economicus* has $\alpha_i = 0$.

In the Modified Dictator game with which β_i is elicited, our findings are comparable with those of Blanco, et al.. Mean values of around 0.5 seem quite high (especially relative to *homo economicus*), but the discussion of Blanco, et al. (p. 333) suggests a possible reason for this shared finding. The active player may feel responsible for the passive player in the Modified Dictator game; and looking out for that player's interests would tend to boost the elicited value of β_i, even for a subject who would not put much weight on the payoff of another in different circumstances.

These arguments suggest that, taken on its own, a finding that some individual violates $\alpha_i \geq \beta_i$ may not be all that surprising, when one keeps in mind that the condition is on revealed preference. Nevertheless, we find more frequent violations than Blanco, et al. had, and this was contrary to our expectations. Further studies would be useful, especially in non-standard subject pools.

That said, the similarities between our findings from distinct subject pools are arguably more striking than the differences, with two exceptions each of which relates to revealed preference. The first is the much greater incidence of non-well-behaved responses to the revealed preference tasks in Izmir than in the

other two subject pools. The second is greater incidence among well-behaved subjects of extreme values (high and low) of α_i among the Izmir subject pool, as compared with Nottingham and MTurk. One possible interpretation of these findings is a lower level of understanding of the relevant tasks in some subject pools. But, we cannot rule out some more fundamental, society-related subject-pool differences (a possibility suggested by Herrmann, et al. 2008).

There is nothing inherently puzzling about one society displaying more extreme values of revealed aversion to disadvantageous inequality than another, especially as this aversion is inferred from the subject's strategy as responder in the Ultimatum game. It may be that, in some societies, there is a strong motivation not to leave money on the table, but this can be over-ridden by a sense of insult and, if it is, then the opposite reaction is also powerful. As Blanco, et al. (Section 7) notes, the Fehr-Schmidt model can be re-interpreted as an indirect reduced-form for reciprocal motivations. Such motivations could affect the aversion to disadvantageous inequality that we infer from the responder's strategy in the Ultimatum game. Thus, a possible explanation of differences between subject pools in this parameter is that they differ either in the strength of their reciprocity or in the consistency across individuals of how they balance reciprocal concerns with pure aversion to inequality.

The interpretation of the Fehr-Schmidt model as a reduced-form for reciprocal motivations is also relevant to points discussed earlier. If the mapping between material inequality and reciprocity is sensitive to context, that might contribute to the weak association which we find between stated and revealed aversion to disadvantageous inequality. To the extent that positive and negative reciprocity are distinct motivations (as is suggested by existing

evidence from related ultimatum and dictator games, e.g., Yamagishi et al., 2012; Peysakhovich, Nowak, & Rand, 2014), this perspective would also help to explain why positive and negative inequality aversion, as revealed in the Blanco, et al. tasks are not strongly positively correlated and weaken any expectation that one will always be stronger than the other.

2.5 Supplementary Materials

2.5.1.1 Further Details on Procedures

2.5.1.1.1 Recruitment

We used ORSEE (Greiner, 2015) to recruit subjects in our Nottingham study. In the UK, students do not typically attend the University closest to their childhood home. By restricting our Nottingham sample to UK citizens, we exclude University of Nottingham students from other countries. In Izmir, recruitment was done by approaching students on campus. This occurred in three primary ways: soliciting volunteers from the end of lectures, contacting participants in the school cafeteria and other social places, and via posters advertising the sessions. Recruitment for the MTurk subjects was done through the creation of two separate 200 subject MTurk HITS (Human Intelligence Tasks).

The way that we structure our MTurk HITS is such that subjects find our HIT on the MTurk platform, click through a link to Qualtrics where they complete the experiment via the Qualtrics survey platform, receive a completion code upon finishing, and then return to MTurk where they input the completion code. Seven participants completed the entirety of the Qualtrics survey, but returned to MTurk after the first 200 participants from their respective MTurk HIT had already entered their completion codes and so were unable to be compensated for their efforts. These subjects contacted the researchers via email and we were able to process their payments via a follow up task.

In the cases where there was an even number of these special participants they were paired together with one other participant from their session and

payment was calculated according to their and their random partners' choices. In the instance of the 407th subject, we randomly selected one of the first 406 and matched that randomly selected subject with the 407th subject calculating payments as if they had actually been matched together. We paid the 407th subject according to what they would have received had they actually been matched with the randomly selected subject. We did not double pay that randomly selected subject from the original 406 subjects.

Some MTurk participants started the experiment, but dropped out before completing the entirety of it. We were able to observe the decisions of all MTurk participants – both those who completed the task and those who did not – via the Qualtrics software. 430 subjects clicked through the link from MTurk and entered the Qualtrics survey. 407 of the 430 (94.65%) subjects completed the entirety of the Qualtrics survey and are included in our analysis. 23 of the 430 (5.35%) subjects clicked through the link from MTurk, entered the Qualtrics survey, and did not complete it in its entirety. 12 of 430 (2.79%) [or 12 of 23 (52.17%) people who started but did not complete the Qualtrics survey] did not even begin the Qualtrics survey. The remaining 11 of 430 (2.56%) [or 11 of 23 (47.83%) people who started but did not complete the Qualtrics survey] all completed the Modified Dictator Game after which 2 more dropped out and only 9 of 430 (2.09%) [or 9 of 23 (39.13%) people who started but did not complete the Qualtrics survey] completed the Ultimatum Game. These 9 eventually dropped out and so we did not include them anywhere in our analysis aside from here.

2.5.1.1.2 Subject Pool Statistics

Table 2.3 includes subject pool details for subjects included in the analysis; that is, subjects who had *well behaved* preferences (as defined in the main text; see also below). The first row gives the proportion of female subjects. Post-Secondary refers to any higher education experience. By default all subjects in the Nottingham and Izmir subject pools meet this requirement while 82% of the MTurk subjects have had some amount of post-secondary education. It has been observed that studying economics or business can impact participants' preferences and for this reason it is included in our analysis. The mean and standard deviation of age is given as well as the minimum and maximum values.

	Nottingham	Izmir	MTurk
Female	46%	41%	46%
Post-Secondary	100%	100%	82%
Economics	15%	16%	1%
Business	7%	14%	15%
Age	20.21	21.51	32.94
	(1.91)	(1.71)	(10.12)
Minimum	18	18	19
Maximum	29	30	75
GASP	4.44	5.05	4.89
	(0.92)	(0.78)	(0.85)
Total Sample Size	104	206	407
Percent Well Behaved	82%	45%	91%

Table 2.3. Subject Pool Details. In the first three sections of the table, subject pool details are given only for subjects whose revealed preferences were well behaved. GASP denotes the score from the guilt and shame proneness scale of Cohen, et al. (2011). A GASP score = 1 (7) indicates low (high) levels of guilt and shame proneness. The fourth section includes information about all participants.

2.5.1.1.3 Order of Tasks

The order of tasks in our experimental sessions was the Blanco et al. (2011) Modified Dictator game and Ultimatum game plus two public goods games (greater details given below) followed by the scenario tasks. This order was chosen to preclude spillover effects from the scenario tasks into the games. We assumed that spill-overs from games to scenarios are less likely than from scenarios to games, but, admittedly, we have not tested this assumption. Each participant made decisions in all the roles of each game using the strategy method where necessary. A game was randomly selected for payment at the end of the experiment (the selection of each game had equal chance, as did each assignment of roles). Participants received no feedback until the end of the whole experiment. In the experiment, we measured earnings in points which were exchanged into local currency at the end of each session.[6]

The public goods games are not analyzed in this paper, but, for sake of completeness, we report what we did. Subjects participated in two two-player one-shot public good games following the same procedures as Fischbacher et al. (Journal of Economic Psychology 33(4), 897-913, 2012). In the first public good game – known as the P-experiment in Fischbacher et al. (2012) – subjects' contribution preferences were elicited using a form of the strategy method. In

[6] In the UK, 1 point = 0.75 pence; in Turkey, 1 point = 1.5 Turkish Lira; and on Amazon Mechanical Turk, 1 point = 17.5 cents. Exchange rates were chosen to equalize purchasing power as much as possible in the student subject pools. Since MTurk is a naturally occurring work-place, a different payment structure (with a higher participation fee) was used, to conform to its conventions.

the second public good game – known as the one-shot C-experiment – subjects' report their belief about their co-player's contribution and then their own contribution.

Subjects did not receive any feedback until after completing the scenario task. The feedback they received was twofold: (1) after all the games and scenario tasks were completed but before the questionnaire subjects learned which of the four games would be payoff relevant (in the Nottingham and Izmir experiments; in the MTurk experiment they received this information after completion of the MTurk HIT) and (2) only after completion of the questionnaire did participants see feedback and at this time they received feedback only on the payoff relevant game (again for the MTurk subjects, they received this information after the completion of the MTurk HIT).

2.5.1.1.4 Assigning Revealed Preferences Values

We also follow Blanco, et al. in assigning $\beta_i = 1$ (resp. $\beta_i = 0$) to subjects who always (resp. never) choose the equal distribution; and in excluding subjects with multiple switch points in the Modified Dictator Game from the analysis. We likewise follow Blanco et al. and assign participants who do not reject any offers an $\alpha_i = 0$ and participants who reject every offer below the equal split, an $\alpha_i = 4.5$ (although in theory these participants could have $\alpha_i \geq 4.5$). Subjects with multiple switch points in the Ultimatum Game are excluded from the analysis in both the main paper and supplementary materials. In summary, subjects were excluded from the analysis by having:

- Multiple switch points in the Modified Dictator Game alone
- Multiple switch points in the Ultimatum Game alone

- Multiple switch points in both the Modified Dictator and Ultimatum Games

2.5.1.2Full Instructions of the Experiment

2.5.1.2.1 Background

Included here are the full instructions for the Nottingham subject pool. These instructions were based off of those by Loewenstein et al. (1989) for the scenarios and Blanco et al. (2011) for the Modified Dictator game and Ultimatum game. Blanco et al. (2011) kindly provided instructions and copies of the zTree files which were used in this experiment with only minor modifications. The text and all the materials including the zTree files for the Izmir subject pool were translated (both forward and reverse) into Turkish. In both Nottingham and Izmir, the experiments were conducted by native speakers and supervised by one of the authors [Beranek]. The Turkish version of scenarios and instructions, as well as all zTree files are available upon request.

The MTurk participants completed the experiment online using the online survey software Qualtrics via MTurk and the text and materials were Americanized. These materials are available upon request in the form of a PDF file.

2.5.1.2.2 Instructions

Economic Research Project

You are now taking part in an experiment. If you read the following instructions carefully, you can, depending on your and other participants' decisions, earn a considerable amount of money. It is therefore important that you take your time to understand the instructions. Please do not communicate with the other participants during the experiment. Should you have any questions, please ask us.

The experiment consists of four different sections. In each section you will be called to make one or more decisions. You will have to make your decisions without knowing other participants' decisions in the previous sections. Note further that the other participants will not know your decisions either.

Only one of the sections will be taken into account in determining your final payoff. This will be randomly determined as described below. Each section has the same probability of being selected. You should take your time to make your decision. All the information you provide will be treated anonymously.

The section that will be taken into account in determining your final payment will be selected as follows. **Participant Number 2** was randomly chosen at the very beginning of the experiment. This participant will draw a ball from a cloth bag **after** all participants have completed all sections. Each ball in the cloth bag has a different colour and each colour corresponds to a different section: yellow, blue, green, and red. The resulting colour and corresponding section will be used to calculate your payment.

The computer will randomly pair you with another participant in the room and will assign the roles. The matching and roles assignment will remain anonymous. You will not know which role you were playing until the end of the game.

Your earnings will be paid to you in cash at the end of the experiment at a rate of 1 point = 50 pence. Earnings will be confidential.

Yellow Section

In this section the situation is as follows:

> *Person A is asked to choose between two possible distributions of money between her and Person B in twenty-one different decision problems. Person B knows that A has been called to make those decisions, and there is nothing he can do but accept them.*
>
> *The roles of Person A and Person B will be randomly determined at the end and will remain anonymous*

Before making your decisions please read carefully the following paragraphs.

The decision problems will be presented in a chart. Each decision problem will look like the following:

Person A's Payoff	Person B's Payoff	Decision	Person A's Payoff	Person B's Payoff
20	0	Left Right	5	5

You will have to decide as Person A; hence if in this particular decision problem you choose left, you decide to keep the 20 points for yourself so Person B's payoff will be 0 points. Similarly, if you choose Right, you and the Person B will earn 5 points each.

You will need to choose one distribution (Left or Right) in each of the twenty-one rows you will have in the screen. If this is chosen as the payoff relevant section, the computer will randomly choose one of the twenty-one decisions. The outcome in the chosen decision will then determine your earnings.

The computer will randomly pair you with another participant in the room and will assign the roles. The matching and roles assignment will remain anonymous.

Please note that you will make all decisions as Person A but the computer might assign you Person B's role.

If you are assigned the role of A, you will earn the amount that you have chosen for Person A in the relevant situation and the person paired with you will earn the amount that you have chosen for Person B.

In the case that you are assigned the role of Person B, you will earn the amount that Person A whom you are paired with has chosen for Person B in the relevant situation.

Blue Section

In this section the situations is as follows:

> *Person A is asked to choose one out of twenty-one possible distributions of money between her and Person B. Person B knows that A has been called to make these decision, and may either accept the distribution chosen by A, or reject it.*
>
> *In the case that Person B accepts A's proposed distribution, that will be implemented. If B rejects the offer, both receive nothing.*
>
> *The roles of Person A and Person B will be randomly determined by the computer and will remain anonymous.*

Before making your decision please read carefully the following paragraphs.

In the case that this section is selected to determine your earnings, the computer will randomly pair you with another participant in the room and will assign the roles. The matching and roles assignment will remain anonymous.

You will have to make decisions as if you were Person A and also as if you were Person B. In the latter case, you will have to decide whether you accept or reject each of A's possible twenty-one proposed distributions.

If you are assigned the role of Person A you will earn the payoff you chose for yourself if the Person B that you are paired with accepts your offer. Otherwise, you both will earn nothing.

If you are assigned the role of Person B, you will earn the payoff that the Person A that you are paired with chose for B, only if you had accepted that particular offer. Otherwise, you both earn nothing.

Green and Red Sections

[These sections were unrelated to this paper.]

Scenarios and Questionnaire

Scenarios

In this section of the project you will read two different scenarios. After reading each scenario, you will learn the outcome of the situation with a variety of

payoffs for you and another party. Your task in this section is to rank your satisfaction with the various payoffs to yourself and the other party on a scale from very unsatisfied (-5) to very satisfied (5). Keep in mind that the order of the payoffs is randomly displayed, so you should be certain to rank your satisfaction of each outcome according to its corresponding payoffs listed to the left of the radio button input scale.

Questionnaire

While calculating your payoff, we would like to ask you to answer the following questionnaire.

Please answer each of the following questions as accurately as possible. Of course, your answers will be treated confidentially. Your honest answers will be of immense value for our scientific investigation. Thank you in advance for your cooperation.

2.5.1.3 Determining Stated Preferences of Inequality Aversion

2.5.1.3.1 Design of Stated Preference Tasks

The scenario tasks provide a near replication of the Loewenstein et al. (1989) (henceforth "LTB89") Study Two scenario tasks with a few exceptions noted below. Participants are asked to rate their satisfaction for outcomes to two scenario disputes. Not every participant faces the same scenarios; dispute type and relationship condition vary across the treatments. The disputes are regarding the gains or losses from disputes involving an invention and from the mutual ownership of a plot of land.

In the original LTB89 paper, the invention scenario regarded the development of cross-country water skis. We developed an alternative invention scenario regarding the development of a smartphone application which is identical in structure to the 1989 scenario, but we expect to be more readily comprehensible to our subjects.

The relationship condition is either a positive or a negative condition and is elaborated in the scenario descriptions. In the MTurk sample, we also included a third condition where there was no relationship manipulation; that is, the nature of the relationship was not mentioned. We refer to this condition as neutral.

We made small adaptations to the scenario text to reflect the individual characteristics of the subject pools – we have an Anglicized version for the Nottingham subject pool, an Americanized version for the MTurk subject pool, and a Turkish version for the Izmir subject pool. The text and all the materials of the Izmir subject pool were translated (both forward and reverse) into Turkish. In both Nottingham and Izmir, the experiments were conducted by native speakers. The scenario text for the Nottingham subject pool is included below. Complete scenario text for the other two subject pools are available upon request.

This was a 2x2x2 design and participants were randomly assigned to each dispute and relationship condition in such a way that they rated both gain and loss conditions for either (a) the invention dispute with a positive relationship condition and the plot dispute with a negative relationship condition or (b) the invention dispute with a negative relationship condition and the plot dispute with a positive relationship condition. For each of the four scenarios, the task is to rate 21 distributions of payoffs for the subject and another person described in the scenario. Each subject is presented with four (out of the eight) scenarios and therefore asked for a total of 84 ratings on a scale from -5 representing “very unsatisfied” to 5 representing “very satisfied.”

The gain conditions are classified as 300, 500, and 600 received to self while the positive outcomes to the other player range from 0 to 900. The loss conditions are the same unit amounts expressed as amounts to pay and not profit. Following the procedures outlined by LTB89, the outcome pairs are randomly ordered to avoid automatic responding. The zTree screen shots from the invention dispute as presented to the Nottingham subject pool are included below in Figure 2.5.

We also included a neutral relationship condition in the American MTurk sample to see what extent the relationship frame impacted utility ratings. In those cases, no relationship information was given to the participants. In this case, this was a 2x3x2 design and participants were randomly assigned to dispute and relationship conditions.

Dispute	Invention		Plot		Invention	Plot
Relationship	Positive	Negative	Positive	Negative	Neutral	Neutral
Gain (21 rankings)	Sequence 1A	Sequence 2A	Sequence 2C	Sequence 1C	Sequence 3A	Sequence 3C
Loss (21 rankings)	Sequence 1B	Sequence 2B	Sequence 2D	Sequence 1D	Sequence 3B	Sequence 3D
BCG_Data file – Scen. Est. tab	Scen1PosDiff Scen1NegDiff	Scen2PosDiff Scen2NegDiff	Scen3PosDiff Scen3NegDiff	Scen4PosDiff Scen4NegDiff	Scen5PosDiff Scen5NegDiff	Scen6PosDiff Scen6NegDiff

Table 2.4. Summary Scenario Sequences and Resulting OLS Estimates.

Table 2.4 summarises the sequences detailing the dispute and relationship conditions present in each as well as referencing the output from the OLS estimations which are included in the Scenario Estimates tab of the BCG_Data file.

- Subjects participated in one of three sequences. In each, they first read the Invention Dispute and then ranked their satisfaction with 21 gain distributions (either Sequence 1A, 2A, or 3A) and then 21 loss distributions (either Sequence 1B, 2B, or 3B).

- Next, they read the Plot Dispute and then ranked their satisfaction with 21 gain distributions (either Sequence 1C, 2C, or 3C) and then 21 loss distributions (either Sequence 1D, 2D, or 3D).
- The sequences varied according to relationship condition: Sequence 1 had a positive relationship frame for the invention dispute and a negative relationship frame for the plot dispute; Sequence 2 had a negative relationship frame for the invention dispute and a positive relationship frame for the plot dispute; and Sequence 3 had neutral relationship frames for both (only half the MTurk participants participated in sequence 3).
- We used each of these rankings as the dependent variable in an OLS estimation for the functional form below with the independent variables being own payoff and the difference between own and other payoff. Each of the sequences resulted in four different OLS parameter estimates (two for NegDiff and two for PosDiff) that can be found in the Scenario Estimates tab of the BCG_Data file which is available as a supplementary file in Excel format.
- The two NegDiff (PosDiff) estimates are averaged together between the scenarios in order to create the Stated *a* (Stated *b*) variables (see discussion in Section 2.5.1.3.3 – particularly page 67 – for further explanation and Table 2.5 for evidence supporting this procedure).

2.5.1.3.2 Scenario Texts

The scenarios were structured in the following way: first, the dispute is introduced; second, a relationship condition is introduced – positive or negative (or neutral in the MTurk subject pool); third, subjects rank their satisfactions with 21 gain distributions and 21 loss distributions. Participants do this entire sequence with two different scenarios for a total of 84 satisfaction ratings.

1. ***Smartphone App Scenario (Updated 2012 version of the 1989 Patent Scenario) with Moderate Relationships, Anglicized***

   a. Dispute: "One day while eating lunch, a student who lives in your residence hall, Charlotte, mentioned to you an idea for a new Smartphone app: a classroom note application for your smartphone. It is similar to a normal word processing app except that you can record lectures, draw diagrams, and take photographs of PowerPoint slides all in real time. Charlotte thought of the idea several years ago, but had not done anything with it and had not been able to interest anyone in it. You find the idea of a classroom note Smartphone app exciting. You suggest to Charlotte that the two of you work together on the project. Over the next month you spend long hours together constructing a prototype of the classroom note app in the computer room. Since it was Charlotte's idea, you agree to pay the rent for the computer room space while you make the app. After extensively testing and refining the classroom note app at your university, you decide that you are ready to submit the app to the Smartphone app store. You complete the Smartphone app store submission, pay the registration fee, and send the app in for approval."

   b. Relationship:

      i. *Moderate positive relationship:* "Charlotte is a student in your residence hall. You like Charlotte a lot, and other people in the dorm also consider Charlotte to be very nice. Charlotte takes notes and picks up assignments for people who miss classes.

Last week, Charlotte made all the arrangements for a small hall party and offered her room to you and your out-of-town guest while she was away over the weekend. In short, Charlotte is kind, friendly, and dependable."

ii. *Moderate negative relationship:* "Charlotte is a student who lives in your residence hall. You have had several unpleasant personal experiences with Charlotte, and other people in the hall also consider Charlotte to be quite rude. Charlotte borrows notes and copies assignments, but does not say thank you and often fails to return items. Last week, Charlotte did not show up for an important intra-mural tournament game and insulted one of your friends. In short, Charlotte is selfish, irresponsible, and argumentative."

c. Outcome:

i. *Gain:* "Several weeks after you submitted the classroom note app to the Smartphone app store, you learn that your Smartphone app has not been approved because there are already similar apps that do the same thing. However, the app store has contacted the developer of one of these similar apps and she is interested in buying one of the innovative features incorporated in your design. You and Charlotte agree that the amount offered seems reasonable. The two of you negotiate how to split the profit."

ii. *Loss:* "Several weeks after you submitted the classroom note app to the Smartphone app store, you learn that your

Smartphone app has not been approved because there are already similar apps that do the same thing. Nevertheless, you are responsible for paying for the Smartphone app store registration fees. Both you and Charlotte receive copies of this bill and negotiate how to split the cost."

2. ***Plot Scenario, with Moderate Relationships, Anglicized***

   a. Dispute: "You live adjacent to an empty plot separating you from your next-door neighbours to your left. No one knows who owns the plot, despite the fact that you and your next-door neighbours have lived there for more than 2 years. However, the local council recently informed you that the plot actually belongs to both you and your neighbours, but the percentage owned by each of you has to be negotiated."

   b. Relationship:

      i. *Moderate positive relationship:* "The Smiths are your neighbours. You like the Smiths a lot, and other neighbours consider the Smiths to be very kind as well. The Smiths always are available to help others. The Smiths are more than happy to water plants and take delivery of parcels when you're away. Last week, the Smiths loaned you some very expensive tools for a repair project and offered their guest bedroom for one of your out-of-town guests. In short, the Smiths are kind, friendly, and dependable."

      ii. *Moderate negative relationship:* "The Smiths are your neighbours. You have had several unpleasant experiences with

the Smiths. Your other neighbours also consider the Smiths to be quite rude. The Smiths borrow things like tools and dishes, but they do not say thank you and often fail to return items. Last week, the Smiths blocked your driveway with their car and threatened to call the police on a small party you were having. In short, the Smiths are selfish, irresponsible, and argumentative."

c. Outcome:

i. *Gain:* "A third neighbour who lacks a garden has agreed to buy the plot. You and your neighbours would both be happy to have a garden between your houses. You and your neighbours need to decide how to split the profits."

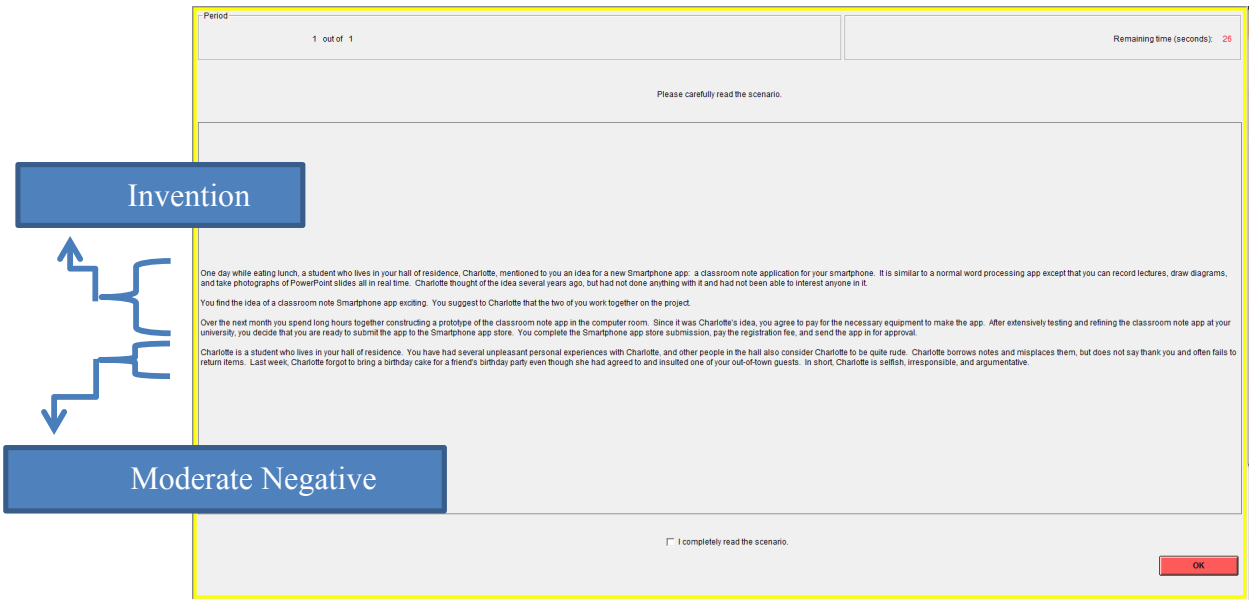
ii. *Loss:* "The plot is too small to sell. However, the local council has amassed taxes on the property that you and your neighbours must pay. You and your neighbours need to decide how to split the costs of the taxes."

The following Figure 2.5 provides an example of the z-Tree screen shots subjects saw in Sequence 2. The screen shots of the other sequences are identical except for relevant differences in text.

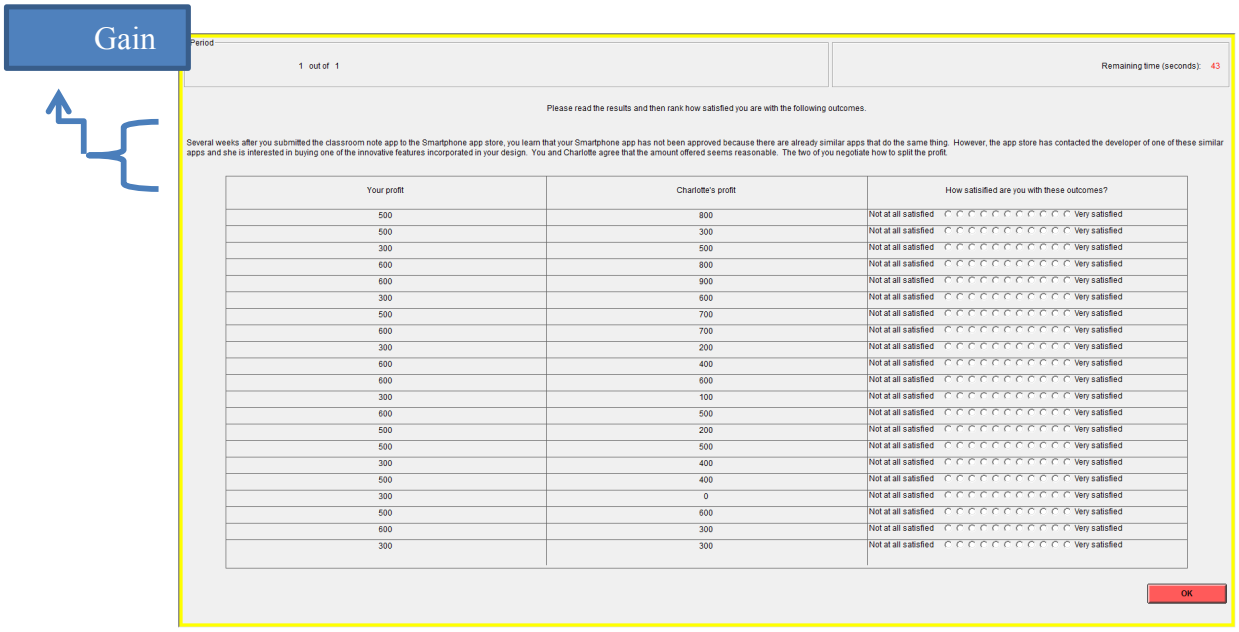
Figure 2.5. Example Screen shots of zTree Scenario Decision Screens

Sequence Two

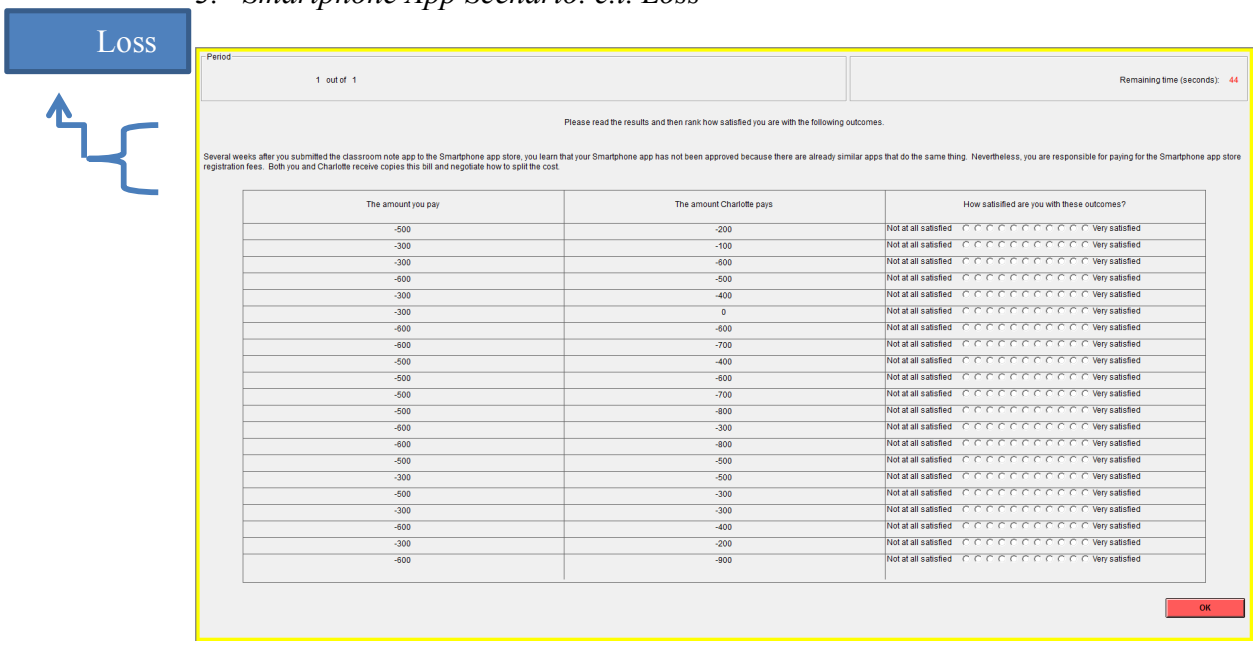
1. *Smartphone App Scenario. a. Dispute and b.ii. Mod. Neg. relationship*



2. *Smartphone App Scenario. c.i. Gain*



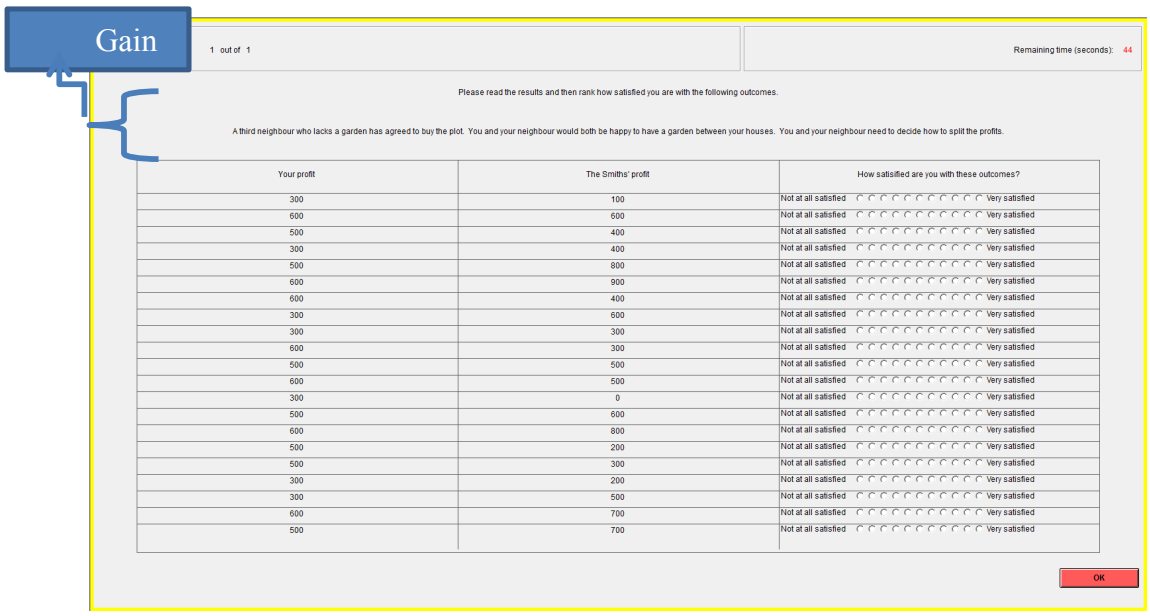
3. *Smartphone App Scenario. c.i. Loss*



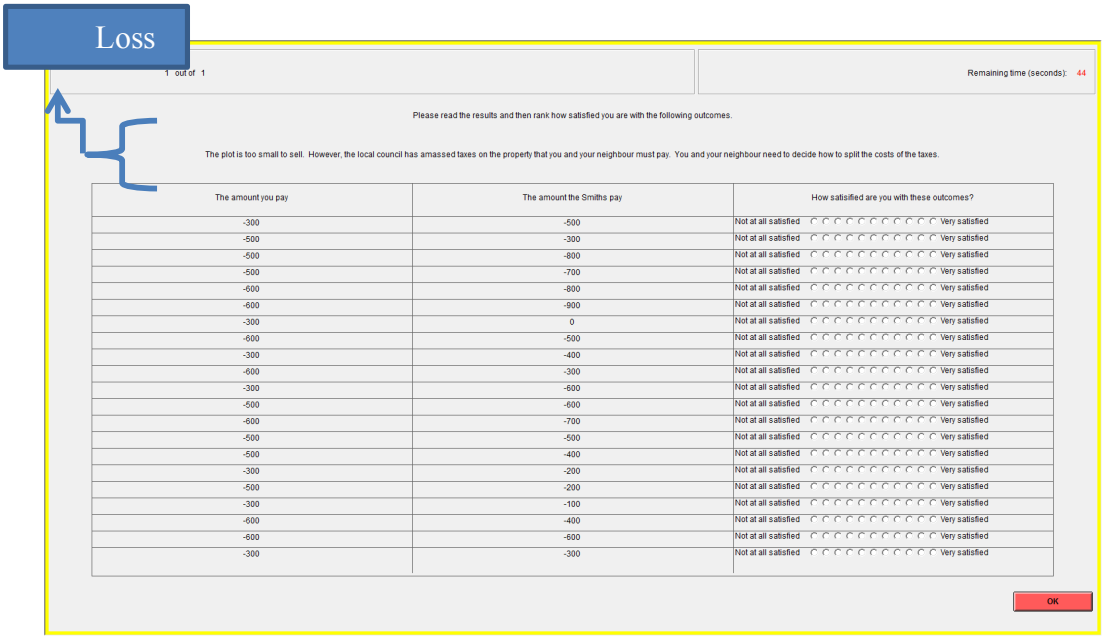
4. *Plot Scenario. a. Dispute and b.i. Moderate positive relationship*



5. *Plot Scenario. c.i. Gain*



6. *Plot Scenario. c.ii. Los*



2.5.1.3.3 Estimation of a_i and b_i (the stated advantageous and disadvantageous inequality aversion parameters)

LTB89 specified their model according to three criteria: goodness of fit, simplicity, and flexibility. They examined five functional forms and the functional form that best satisfied their specifying criteria was one that included payoff for self and relative payoffs (positive and negative differences between own and other payoffs and their squared terms):

$$U_i = c_i + B_1 Self + B_2 NegDiff + B_3 NegDiff^2 + B_4 PosDiff + B_5 PosDiff^2$$

It should be noted that this functional form is quadratic which stands in contrast to the piece-wise linear nature of the Fehr-Schmidt model.

LTB89 used a subject's satisfaction ratings of each distribution as the dependent variable and then performed an OLS estimation to determine the parameters of the same model for each individual subject. Their study includes analysis both on an individual and on an aggregate level. LTB89 averaged together all of the individual level parameter estimates in order to make aggregate level figures and general statements.

For each of our subject pools, we followed LTB89's procedures and also tested each of the five functional forms suggested by LTB89. The same functional form (listed above) – as was the case in LTB89 – had the best goodness of fit as expressed by highest adjusted R^2 values for all of our subject pools.

One of the most familiar figures from the LTB89 paper is shown in Figure 2.6. The notable features of this figure are the tent like structure where highest utility is expressed when both payoffs are equal. Unequal payoffs lead to

decreases in utility, but these decreases in utility are not equal. Utility is reduced more in the region of disadvantageous inequality than in the region of advantageous inequality. These are the features that Fehr and Schmidt (1999) cite as inspiration in the development of different aspects of their model.

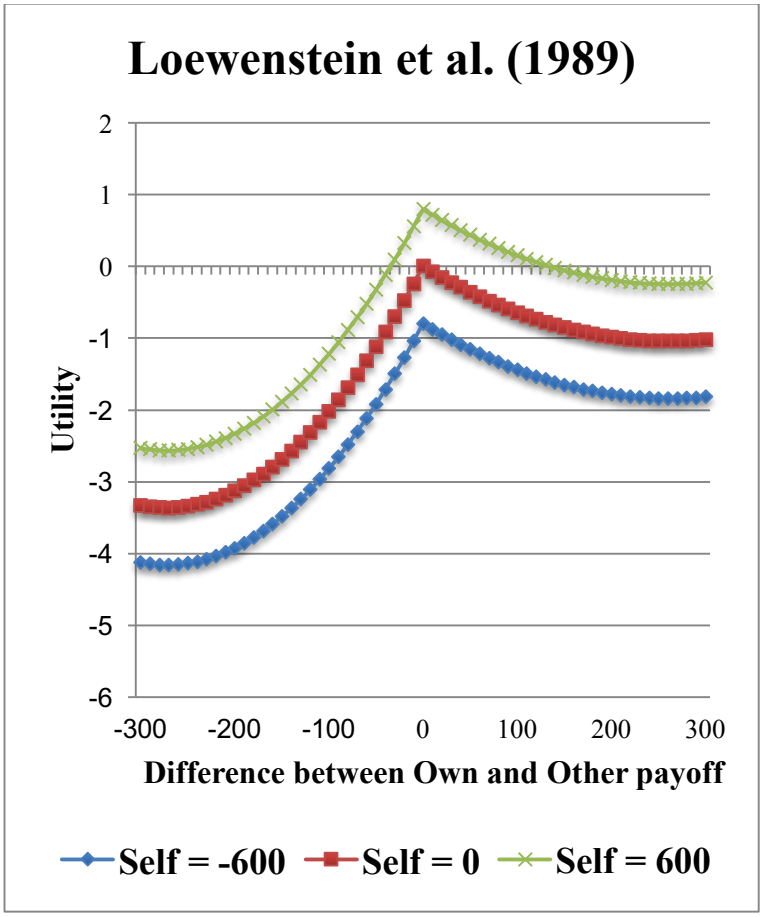


Figure 2.6. The original quadratic LBT89 functional form expressed as a social utility curve emphasizing the importance of relative payoffs.

However, the Fehr and Schmidt parameters of inequality aversion are not directly comparable to the mean parameter estimates in the LTB89 model or the social utility curve shown here in Figure 2.6. The reason why direct comparison is not possible is because the functional form that LTB89 adopt is piece-wise quadratic whereas the model suggested by Fehr and Schmidt (1999) is piece-wise linear.

In order to make direct comparisons between the stated preferences as expressed in the scenarios and the revealed preferences elicited in the Blanco et al. (2011) games, we need to specify a piecewise linear functional form that emphasizes the importance of payoff differences. We can use the satisfaction ratings of each distribution as our dependent variable and then perform an OLS estimation to determine the parameters of a piece-wise linear model which is directly comparable to the Fehr and Schmidt (1999) model:

$$U_i = c_i + x_i - a_i \max\{x_j - x_i, 0\} - b_i \max\{x_i - x_j, 0\}, j \neq i \qquad 2.4$$

We estimate these parameter values for both of the scenarios each participant considers and then average the parameters together. The resulting averages are what we refer to as each individual's stated preferences of inequality aversion (a_i and b_i).

Some might find this averaging procedure questionable as both the dispute and relationship conditions vary between the two scenarios participants consider. A simple way to test this would be to look at the individual level correlation of the estimated parameters across the two scenarios each participant considered. Significantly high and positive Spearman rho values would give support to this technique. Table 2.5 shows the summary statistics and the correlations between the scenarios subjects considered. In all cases but one (that is, in 13 out of 14 cases), there is strong and significant positive correlation between the parameter estimates resulting from the satisfaction ratings in the two separate scenarios considered by each participant. In light of this, we follow LTB89 to average each individual's two parameter estimates to come up with our stated preferences of inequality aversion (a_i and b_i).

a (negdiff) Correlation between Scenarios							
		Nottingham		Izmir		MTurk	
Relationship	*Dispute Type*	*Obs*	*Mean*	*Obs*	*Mean*	*Obs*	*Mean*
Positive	Invention	42	0.0102	38	0.0062	90	0.0113
Negative	Invention	43	0.0127	53	0.0103	93	0.0149
Positive	Plot	43	0.0148	53	0.0116	93	0.0155
Negative	Plot	42	0.0155	38	0.0126	90	0.0179
Neutral	Invention					189	0.0125
Neutral	Plot					189	0.0176
Spearman Correlation		*ρ*	*p*	*ρ*	*p*	*ρ*	*P*
Pos Invent, Neg Plot		0.6252	0.0000	0.4718	0.0028	0.4909	0.0000
Nega Invent, Pos Plot		0.6412	0.0000	0.5566	0.0000	0.4748	0.0000
Neutr Invent, Neutr Lot						0.6109	0.0000

b (posdiff) Correlation between Scenarios							
		Nottingham		Izmir		MTurk	
Relationship	*Dispute Type*	*Obs*	*Mean*	*Obs*	*Mean*	*Obs*	*Mean*
Positive	Invention	42	0.00693	38	0.00826	90	0.00981
Negative	Invention	43	0.00050	53	0.00334	93	0.00363
Positive	Plot	43	0.00947	53	0.00964	93	0.01216
Negative	Plot	42	0.00010	38	-0.00343	90	-0.00136
Neutral	Invention					189	0.00824
Neutral	Plot					189	0.00744
Spearman Correlation		*ρ*	*p*	*ρ*	*p*	*ρ*	*P*
Pos Invent, Neg Plot		0.4105	0.0069	***0.2199***	***0.1845***	0.3069	0.0033
Nega Invent, Pos Plot		0.6283	0.0000	0.6309	0.0000	0.4337	0.0000
Neutr Invent, Neutr Lot						0.6256	0.0000

Table 2.5. Summary statistics and correlations between Loewenstein et al. Dispute Conditions (Participants with Well Behaved Preferences).

Figure 2.7 shows the estimated linear social utility curves for each subject pool. In these figures, the constants are normalized to zero. In these curves, utility is a function of the difference between Own and Other payoff at various levels of one's own payoff (-600, 0, 600). Figure 2.1A in the main text is constructed from these same parameter estimates in the condition when one's own payoff is zero; that is, Figure 2.1A is constructed from the middle series from each of the four subject pool graphs in Figure 2.7. Readers will notice that, aside from their linearity, the Figure 2.7 social utility curves appear similar to

the one in Figure 2.6 (utility is highest for equal payoffs and disadvantageous inequality is disliked more than advantageous inequality).

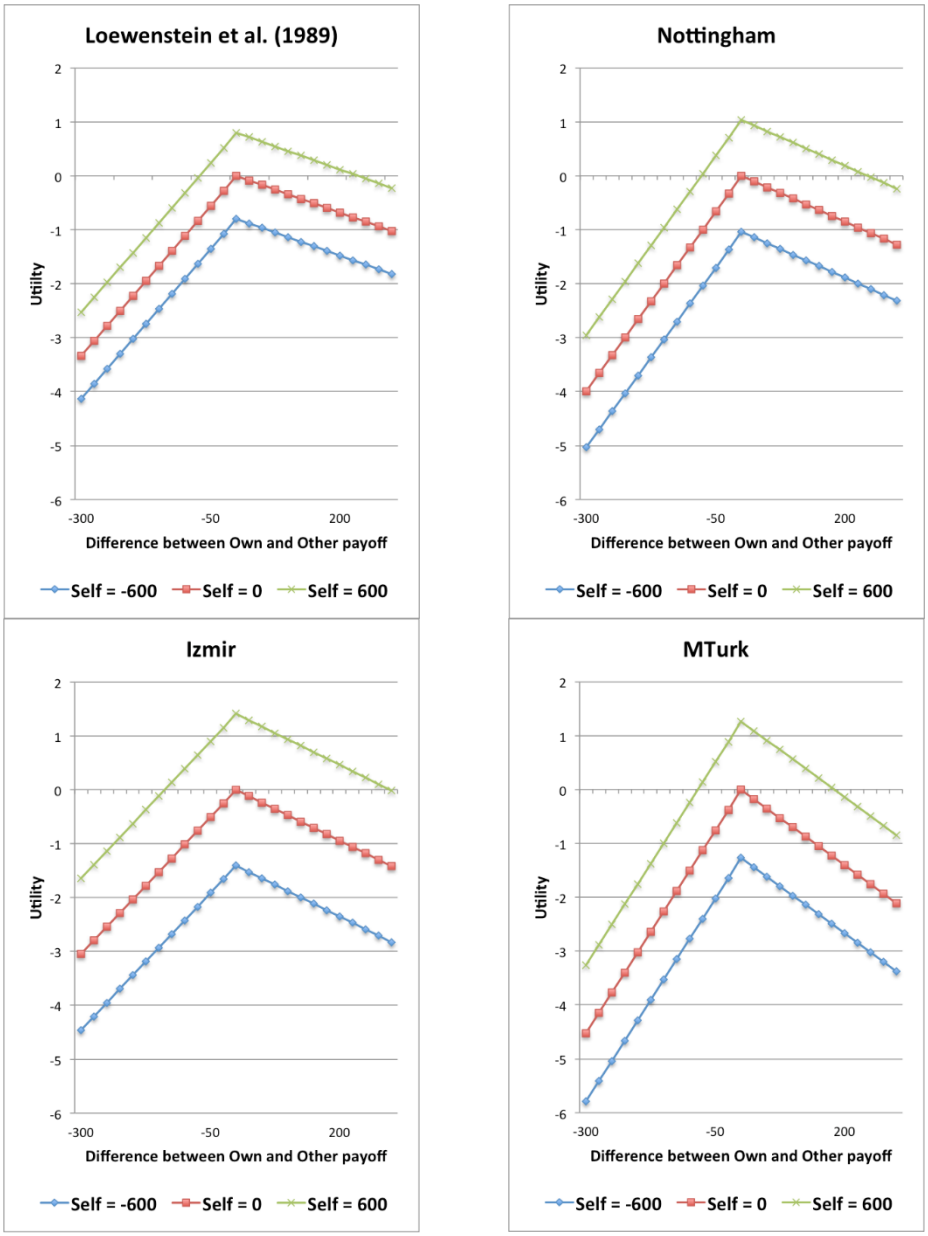


Figure 2.7. Linear Social Utility Curves.

LTB89 did not specify a piece-wise linear model and since their individual level data are not available we are unable to do so ourselves. We instead constructed a linear social utility curve for LTB89 by separating the social utility curve into two components: one component in the domain of advantageous inequality and one component in the domain of disadvantageous

inequality. We then fitted a line to the curve in each component. We used the slope of this fitted line as the average parameter estimate for the a parameter and the b parameter. As a check for the validity of this approach, we followed the same procedure for our Nottingham subject pool (constructing a and b values from the quadratic model that we estimated) and we report both the constructed and estimated the values for a and b in Table 2.6. The differences between our constructed and estimated Nottingham parameters are not big and therefore we deem this an appropriate approximation given the data limitations.

	a	b
LBT89 Constructed	0.01111	0.00341
Nottingham Constructed	0.01334	0.00558
Nottingham Estimated	0.01331	0.00426
Izmir Estimated	0.01018	0.00473
MTurk Estimated	0.01509	0.00703

Table 2.6. Comparing the constructed parameters to the estimated parameters.

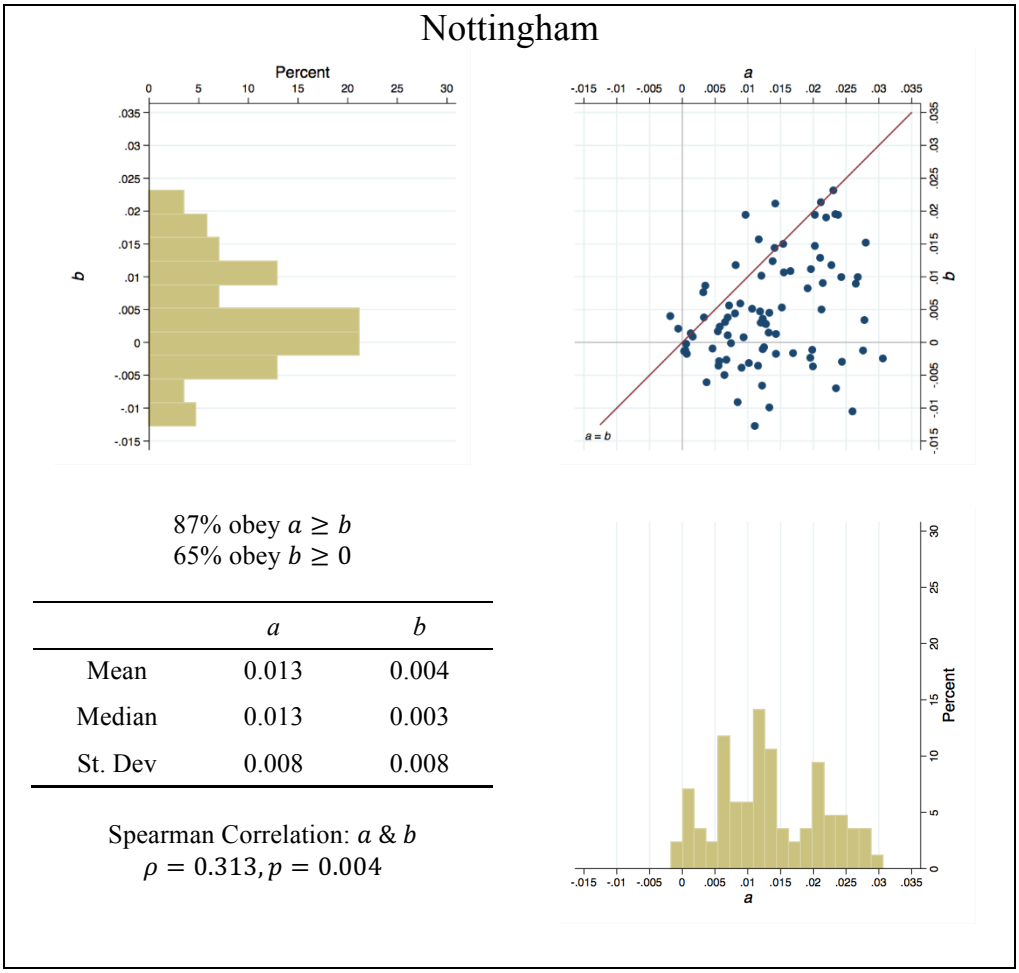
The Kruskal-Wallis tests below in Table 2.7 and the Wilcoxon-Mann-Whitney tests in Table 2.8 tell us whether the differences in distributions of the parameters of inequality aversion (including the variation in stated preferences) by populations are significant.

2.5.1.4 Supporting Analysis

The data reported here are available as a supplementary file (BCG_Data.xlsx).

2.5.1.4.1 Supporting Analysis for Section 2.3.1 Stated Inequality Aversion

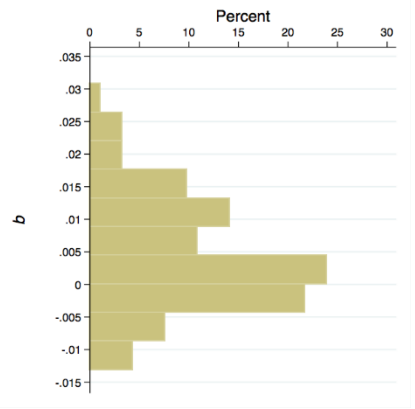
Figure 2.8.[7] **Expanded Versions of Figure 2.1B to Figure 2.1D in the Main Text.** Joint a and b distributions per subject pool. Each dot represents a participant's a and b parameters as calculated from their stated preferences in the updated Loewenstein, et al (1989) scenario tasks. Observations to the left of the $a = b$ line have $a < b$ which violates the Fehr and Schmidt (1999) assumption. The left and bottom panel are histograms of the b and a values, respectively.

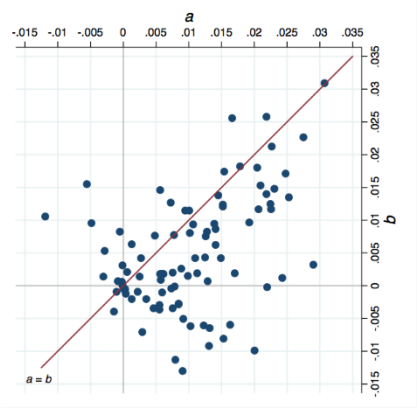


	a	b
Mean	0.013	0.004
Median	0.013	0.003
St. Dev	0.008	0.008

[7] In order to document our elicited a and b values, as well as our α and β values, we show below expanded versions of Figure 2.1B to Figure 2.1D and Figure 2.2. In addition to the scatter plots of the figures in the main text, these expanded figures contain histograms of the distributions of the respective values. This exposition is inspired by Dannenberg et al. (2007).

Izmir

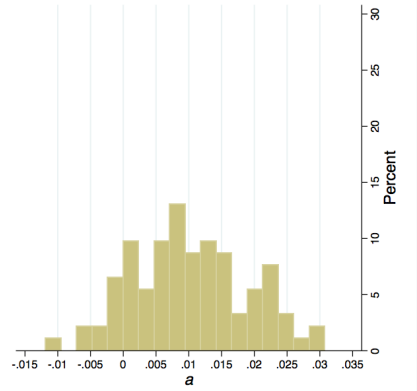




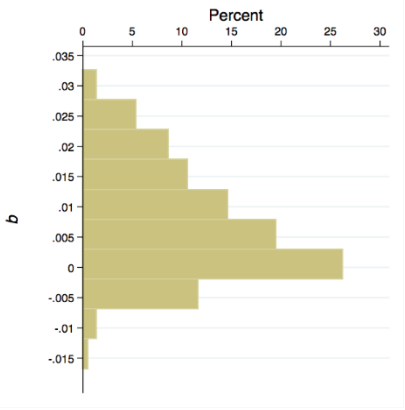
77% obey $a \geq b$
68% obey $b \geq 0$

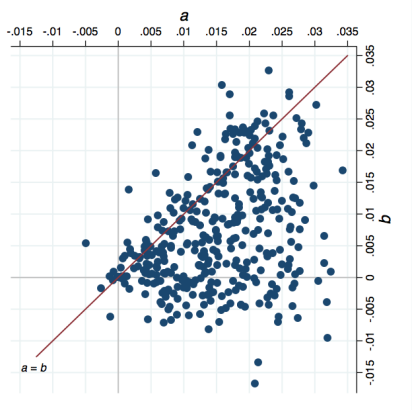
	a	b
Mean	0.010	0.005
Median	0.009	0.002
St. Dev	0.009	0.009

Spearman Correlation: a & b
$\rho = 0.406, p = 0.000$



MTurk

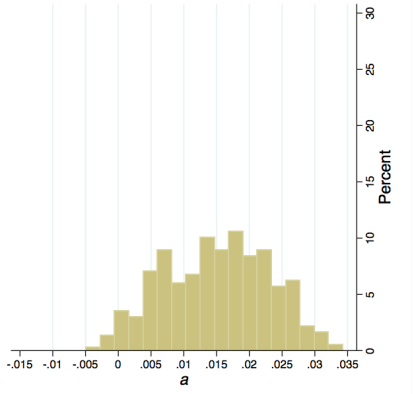




80% obey $a \geq b$
76% obey $b \geq 0$

	a	b
Mean	0.015	0.007
Median	0.015	0.005
St. Dev	0.008	0.009

Spearman Correlation: a & b
$\rho = 0.354, p = 0.000$



We list the results of Kruskal-Wallis tests in Table 2.7 indicating that for the most part the samples do not seem to originate from the same distribution beyond the categorization imposed above.

	χ^2 (2)	*p*	χ^2 (2) with ties	*p*
a (negdiff)	23.858	0.0001	23.858	0.0001
b (posdiff)	8.779	0.0124	8.779	0.0124

Table 2.7. Kruskal-Wallis Tests comparing parameters of stated inequality aversion by populations.

In Table 2.8, we list the results of Wilcoxon-Mann-Whitney tests comparing the parameters of inequality aversion bilaterally between populations. We note that significant differences exists between all populations with regards to the stated *a* values. In regards to the stated *b* values, significant differences exist between MTurk and both Nottingham and Izmir populations.

	Notts MTurk		Notts Izmir		Izmir MTurk	
	Z	*p*	Z	*p*	Z	*p*
a (negdiff)	***1.881***	***0.0600***	***-2.331***	***0.0197***	***4.762***	***0.0000***
b (posdiff)	***2.455***	***0.0141***	0.244	0.8075	***2.107***	***0.0351***

Table 2.8. Wilcoxon-Mann-Whitney test comparing parameters of stated inequaliity aversion between populations.

2.5.1.4.2 Supporting Analysis for Section 2.3.2 Revealed Inequality Aversion

The Chi^2 tests listed in Table 2.1 of the main text compare the distributions of each subject pool to the theoretical Fehr and Schmidt (1999) distribution and the observed Blanco et al. (2011) (BEN) distributions. Table 2.9 shows group Chi^2 tests with different combinations of groupings. The upper portion of the table involves comparisons to the theoretical Fehr and Schmidt (1999) distributions. The lower portion of the table involves comparisons to observed data. Various groupings are considered including combining the Blanco et al.

(2011) observations with our Nottingham observations for a UK university group. For the most part group Chi^2 tests indicate that our groups are significantly different from one another. The exception is greater similarity between groups of observed β distributions as opposed to the comparison of observed β distributions to the Fehr and Schmidt (1999) theoretical distribution.

α Distributions		***β Distributions***	
Comparison to F&S Theoretical Distribution			
Groups: F&S, BEN, Nottingham, Izmir, Mturk			
$\chi 2$ (12)	53.982	$\chi 2$ (8)	27.562
p value	0.000	p value	0.001
Groups: F&S, Nottingham, Izmir, Mturk			
$\chi 2$ (9)	45.288	$\chi 2$ (6)	24.682
p value	0.000	p value	0.000
Groups: F&S, BEN & Nottingham Combined, Izmir, Mturk			
$\chi 2$ (9)	44.798	$\chi 2$ (6)	24.638
p value	0.000	p value	0.000
Comparison to Observations			
Groups: BEN, Nottingham, Izmir, Mturk			
$\chi 2$ (9)	34.484	$\chi 2$ (6)	11.566
p value	0.000	p value	0.072
Groups: Nottingham, Izmir, Mturk			
$\chi 2$ (6)	21.476	$\chi 2$ (4)	7.950
p value	0.002	p value	0.093
Groups: BEN & Nottingham Combined, Izmir, Mturk			
$\chi 2$ (6)	25.176	$\chi 2$ (4)	8.403
p value	0.000	p value	0.078

Table 2.9. Group Chi^2 Tests Comparing α and β Categories.

We list the results of Kruskal-Wallis tests in Table 2.10 indicating that there are weakly significant differences between our subject pools in β values; while, the α values do not seem to be originating from independent distributions. Note that we include the Blanco et al. (2011) data in the first instance and exclude it in the second.

With BEN	χ^2 (3)	*p*	χ^2 (3) with ties	*p*
α	***6.033***	***0.1100***	***6.215***	***0.1016***
β	7.359	0.0613	7.444	0.0590
Without BEN	χ^2 (2)	*p*	χ^2 (2) with ties	*p*
α	***3.308***	***0.1913***	***3.422***	***0.1807***
β	6.492	0.0389	6.575	0.0373

Table 2.10. Kruskal-Wallis Tests comparing parameters of revealed inequality aversion by populations.

The results of Wilcoxon-Mann-Whitney tests comparing the parameters of inequality aversion bilaterally between populations are listed in Table 2.11. The α values between Nottingham and both Izmir and MTurk seem similar, but there are significant differences between Izmir and MTurk. In regards to β values, we note that Izmir is different than both Nottingham and MTurk (whereas Nottingham and MTurk are not significantly different from one another.

	Nottingham MTurk		Nottingham Izmir		Izmir MTurk	
	Z	*p*	Z	*p*	Z	*p*
α	0.591	0.5544	-1.432	0.1522	***1.739***	***0.0821***
β	0.860	0.3898	***2.411***	***0.0159***	***-2.195***	***0.0282***

Table 2.11. Wilcoxon-Mann-Whitney test comparing parameters of revealed inequality aversion between populations.

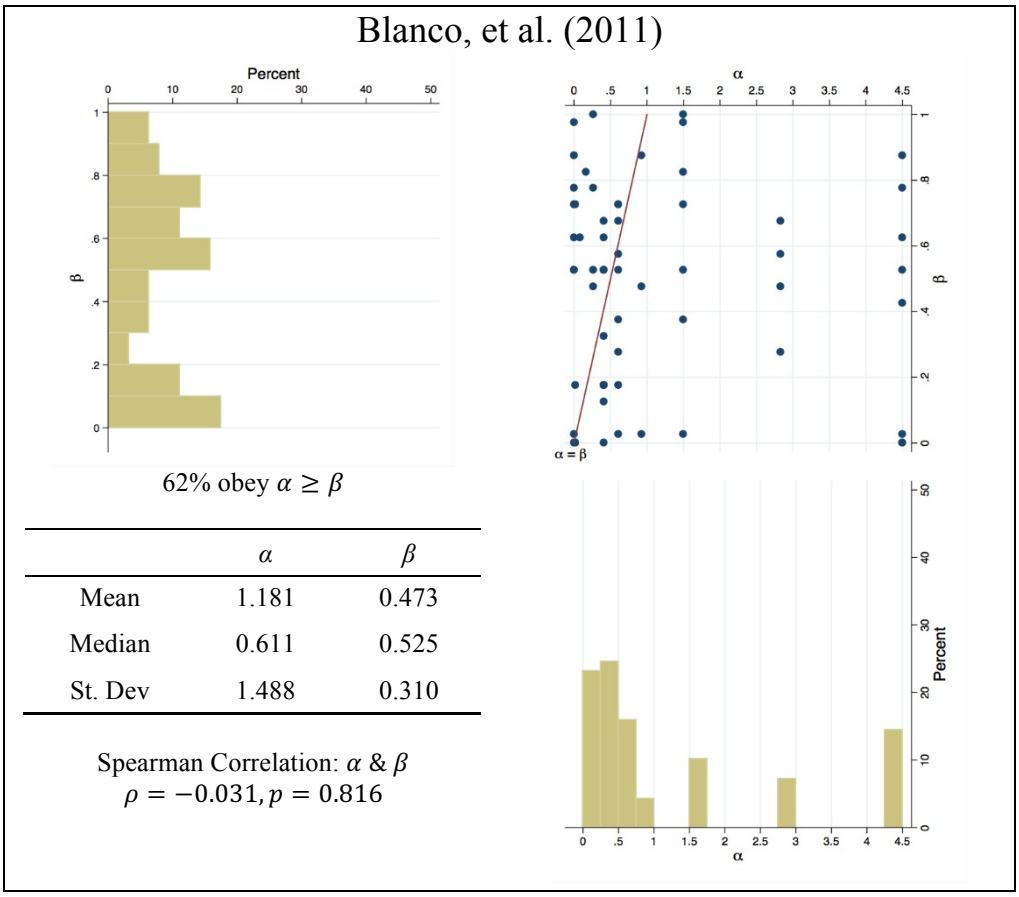
Finally, we note the significant variation in the percent of 'well-behaved' participants in each subject pool with only 45% of our Izmir sample meeting the criteria of well-behavedness as defined in the main text. Several referees requested information about the non well-behaved subjects in Izmir and here we report the proportions excluded for having multiple switch points in the various games in Table 2.12.

		Modified Dictator Game		
		Multiple Switches	*Single Switch*	
Ultimatum Game	*Multiple Switches*	52	17	69
	Single Switch	45	92	137
		97	109	

Table 2.12. Incidences of Multiple Switching in Izmir.

Approximately 25% of subjects in Izmir have multiple switch points for both the MDG and the UG. About 22% have multiple switch points in just the MDG compared to just around 8% who have multiple switch point in just the UG. Again, 45% have well behaved preferences. There seem to be significantly more people who have multiple MDG switches as opposed to UG switches. This pattern is true in Izmir (elsewhere) where approximately 3x (5x) as many people have multiple switches in the MDG compared to the UG. There are also more people who report multiple switch points for both the MDG and the UG in Izmir (approximately 25% of subjects from that subject pool) versus elsewhere (approximately 2%).

Figure 2.9. Expanded Versions of Figure 2.2 in the Main Text. Joint α and β Distributions. Each dot represents a participant's α and β parameters as calculated from their revealed preferences in the Blanco, et al (2011) games. Observations to the left of the $\alpha = \beta$ line have $\alpha < \beta$ which violates the Fehr and Schmidt (1999) assumption. The left and bottom panel are histograms of the β and α values, respectively.

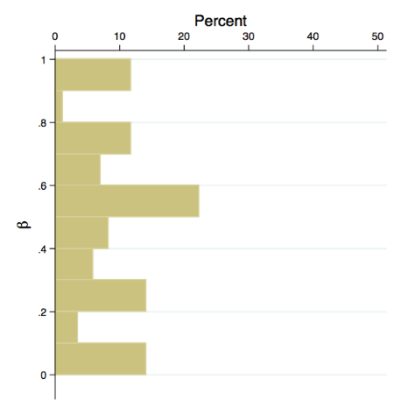


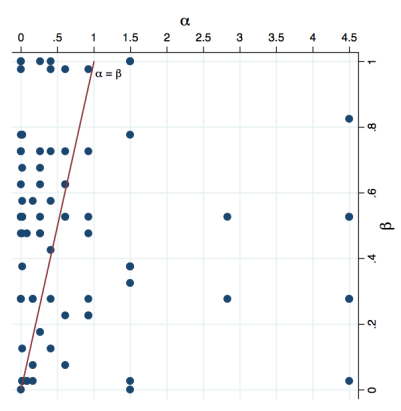
62% obey $\alpha \geq \beta$

	α	β
Mean	1.181	0.473
Median	0.611	0.525
St. Dev	1.488	0.310

Spearman Correlation: α & β
$\rho = -0.031, p = 0.816$

Nottingham

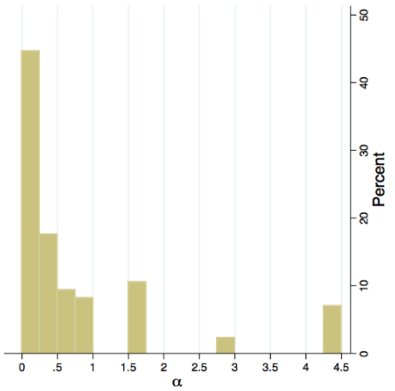




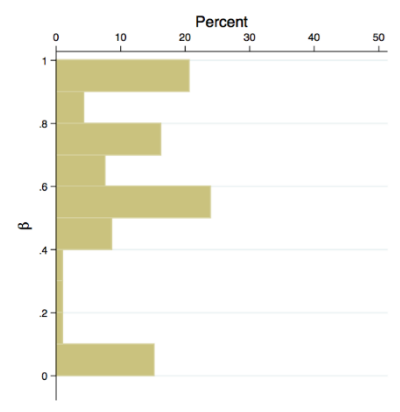
45% obey $\alpha \geq \beta$

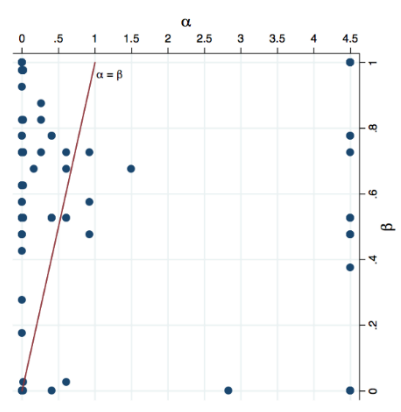
	α	β
Mean	0.754	0.484
Median	0.269	0.525
St. Dev	1.198	0.290

Spearman Correlation: α & β
$\rho = -0.088, p = 0.422$



Izmir

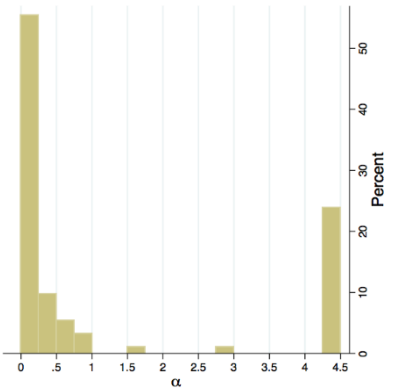




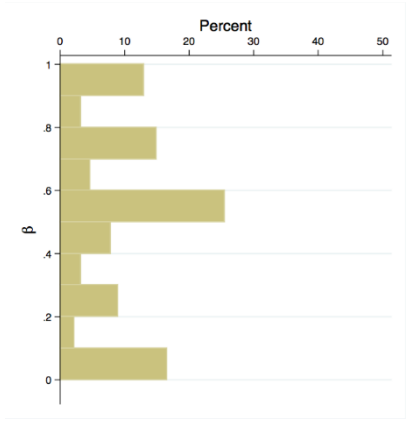
41% obey $\alpha \geq \beta$

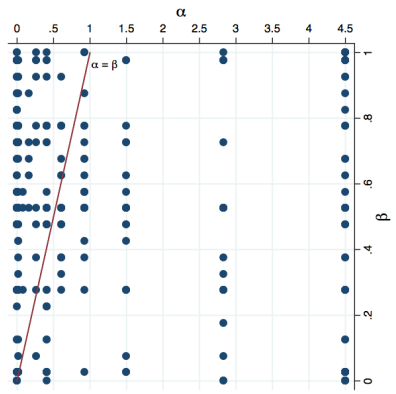
	α	β
Mean	1.227	0.589
Median	0.026	0.575
St. Dev	1.884	0.315

Spearman Correlation: α & β
$\rho = 0.044, p = 0.676$



MTurk

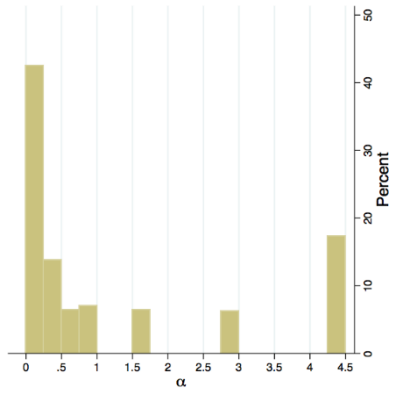




49% obey $\alpha \geq \beta$

	α	β
Mean	1.218	0.512
Median	0.410	0.525
St. Dev	1.670	0.302

Spearman Correlation: α & β
$\rho = -0.107, p = 0.040$



Chapter 3

Does inequality aversion explain free riding and conditional cooperation?[8]

3.1 Introduction

The social dilemma in a public goods game arises when, at the margin, social benefits of contributing exceed the privately borne costs of doing so, but the private benefits do not. As a result, the standard theoretical prediction is under provision of public goods (Samuelson, 1954). However, we observe positive contributions to public goods in both the real world and a large number of experiments (Ledyard, 1995; Chaudhuri, 2011) and, as a result, a common outcome in public goods games is inequality in payoffs among players. People's

[8] This chapter draws on joint work in progress with Robin Cubitt and Simon Gächter.

willingness to contribute at all suggests they have some form of other-regarding preferences. But, the conditionality of willingness suggests an aversion to unilateral contribution. This chapter explores a possible source of that aversion, namely an underlying dislike of being worse off in material terms than those one is interacting with. This is a possible explanation of these contribution patterns because unilateral contribution lowers the material well-being of the contributor relative to other beneficiaries of the public good.

Fischbacher, Gächter, & Fehr (2001) [FGF] introduce a public good game played with the strategy method in which subjects, in an incentive compatible manner, indicate on a contribution table their contribution to the public good for each level of possible coplayer contribution by others. It enables the systematic classification of players into heterogeneous types including free riders who never contribute to the public good regardless of their coplayers' contributions and conditional cooperators who condition their own contributions on those of their coplayers.

While this phenomenon of conditional cooperation has been frequently observed (FGF; Fischbacher & Gächter, 2010; Fischbacher, Gächter, & Quercia, 2012), the behavioral mechanisms by which it operates have not been clearly identified. Some theoretic possibilities have been suggested including betrayal aversion (Bohnet & Zeckhauser, 2004; Cubitt, Gächter, & Quercia, 2015), reciprocity (Dufwenberg, Gächter, & Hennig-Schmidt, 2011), intentions based reciprocity and fairness concerns (Falk & Fischbacher, 2006), altruism (Croson, 2007), and guilt aversion (Charness & Dufwenberg, 2006).

We consider perhaps the simplest possibility: a theory of inequality aversion based solely on payoff comparisons. According to this type of theory players gain utility from their own material payoffs and potentially experience disutility as there is inequality in payoffs between oneself and one's comparator. The Fehr and Schmidt (1999) model of inequity aversion makes the straightforward prediction that a player in a two-player linear public good game will either free ride or match the contribution of their co-player, depending on the strength of their aversion to inequality. A player unconcerned with comparisons of payoffs would focus solely on maximizing her own payoff and therefore would not contribute to the public good at all regardless of her coplayer's contribution. In contrast, a sufficiently inequity averse player will seek to eliminate any differences in payoffs by matching her contributions to those of her coplayer.

Our research strategy to evaluate the predictions of this theory relies on eliciting a parameter of advantageous inequality aversion using the Blanco, Engelmann, & Normann (2011) modified dictator game and determine type using the FGF public good game played with the strategy method.[9] We then look for association between this parameter and a subject's type, measured by the FGF method, paying particular attention to free riding and conditional cooperation. We use two-player public good games because this makes the idea and detection of conditional cooperation particularly straightforward. Previous research suggests we will find a heterogeneity of degrees of inequality aversion

[9] The focus of this chapter is on the revealed preference measure of beta. This remains the standard approach of economics, notwithstanding the previous chapter.

as well as substantial proportions of conditional cooperators and free riders. We then investigate the extent to which inequality aversion explains free riding and conditional cooperation.

In Section 3.2, we describe our research strategy in greater detail and show more specifically how the theory makes predictions about the observed heterogeneous types of players. In Section 3.3, we detail the design of our experiment. Section 3.4 reports our results showing the distribution of and associations between types enabling us to answer our primary research question of whether inequality aversion explains free riding and conditional cooperation. We conclude in Section 3.5.

3.2 Research Strategy

Our research strategy depends on the modelling of inequity averse preferences and the environment of the public good game. We detail these both in Subsections 3.2.1 and 3.2.2, respectively. Combining preferences with the specifics of the environment enables us to make predictions of game play in Section 3.2.3

3.2.1 Preferences

While there are many models of inequality averse preferences including Loewenstein, Thompson, & Bazerman (1989), Bolton (1991), and Bolton & Ockenfels (2000), we focus on perhaps the most popular model of these preferences: the Fehr and Schmidt (1999) theory of inequity aversion. At the center of this theory is an elegant, intuitive, and tractable piece-wise linear model of utility made up of three components: own payoff, disutility from disadvantageous inequality, and disutility from advantageous inequality.

Formally, the two-player version of the Fehr and Schmidt (1999) model of inequity aversion is:

$$U_i = x_i - \alpha_i \max\{x_j - x_i, 0\} - \beta_i \max\{x_i - x_j, 0\},\ j \neq i \qquad 3.1$$

$$\text{where } \beta_i \leq \alpha_i \text{ and } 0 \leq \beta_i < 1$$

In this model, player *i*'s utility is given by U_i and her monetary payoff is x_i. When making comparison between her own and her coplayer's payoff, player *i* can find herself in one of three possible situations: one of equality of payoffs where $x_i = x_j$, one of disadvantageous inequality where $x_i < x_j$, and one of advantageous inequality where $x_i > x_j$. Inequality becomes relevant to one's payoff depending on one's own inequality aversion. The nature of the Fehr and Schmidt (1999) model allows player *i* to evaluate disadvantageous and advantageous inequality separately according to her specific parameters of inequality aversion.

Disadvantageous inequality aversion (the strength of which is governed by the parameter α_i) reduces player *i*'s utility in a piece-wise linear manner dependent on the magnitude of disadvantageous inequality. Likewise, advantageous inequality aversion (the strength of which is governed by the parameter β_i) reduces player *i*'s utility in a piece-wise linear manner dependent on the magnitude of advantageous inequality. The advantage of the two-player structure that we adopt is that a player can only experience one type of inequality at a given moment. Given that these situations are mutually exclusive, at any given time at most only one parameter of inequality aversion is relevant to each of our players. The Fehr and Schmidt (1999) model places further assumptions on the parameters of inequality aversion: (1)

disadvantageous inequality aversion must be greater than advantageous inequality aversion; that is, $\alpha_i \geq \beta_i$, and (2) advantageous inequality aversion must be zero or positive and smaller than unity; that is, $0 \leq \beta_i < 1$.

3.2.2 Environment

The environment of this study is a two-player linear public good game of standard voluntary contribution mechanism (VCM) form. The monetary payoffs of this game are:

$$x_i = E - c_i + m(c_i + c_j) \quad 3.2$$

$$x_j = E - c_j + m(c_i + c_j) \quad 3.3$$

where $c_i,\ c_j \in \{0, \dots, E\}$; $0.5 < m < 1$

At the start of the game, both players are endowed with the same amount of resources (E). Each player chooses how much of her endowment to contribute to the public good (c_i and c_j, for players *i* and *j* respectively) retaining the rest for herself. Payoffs are comprised of two components: the retained portion of the endowment and the return from the public good, which is the sum of the contributions of both players multiplied by the marginal per capita return (m).

The marginal per capita return (m) in a two-player public good game is restricted to values between 0.5 and 1. This restriction creates a social dilemma because it implies $2m > 1 > m$; that is, the marginal social returns of contribution exceed the marginal costs which, in turn, exceed the marginal private returns from contribution. The socially optimal outcome is full contribution by both players. However, the largest possible individual payoff is

realized by choosing to free ride and not contribute irrespective of the contribution decision of one's coplayer.

The implication of this is that player *i*'s payoff is greater [less] than player *j*'s as player *j*'s contribution to the public good is greater [less] than player's *i*'s:

$$x_i \gtrless x_j \; as \; c_j \gtrless c_i. \qquad 3.4$$

Inequality in monetary payoffs exists when the players contribute different amounts to the public good. The player who contributes less to the public good is better off in monetary terms. Equal contributions eliminate this inequality.

3.2.3 Predictions

Combining the preferences from the Fehr and Schmidt (1999) theory of inequity aversion with the environment of the two-player public good game enables us to make predictions in terms of a player's optimal contribution, for any given level of player *j*'s contribution c_j. First consider a case where player *i* contributes more than player *j*, i.e. $c_i > c_j$. In this case, player *i*'s utility is reduced in two ways, relative to the situation where $c_i = c_j$. First, player *i*'s utility is reduced because she contributes more, thereby reducing her own monetary payoff as $m < 1$. Second, contributing more than her coplayer increases the amount of disutility she experiences due to disadvantageous inequality aversion. In both of these ways, player *i*'s utility is reduced, relative to the situation where $c_i = c_j$. Consequently, player *i*'s optimal choice will never be to contribute more than player *j*.

Now imagine the case where player *j* contributes more than player *i*. The piece-wise linearity of the Fehr and Schmidt (1999) utility function means that (apart from in the case of indifference), player *i* always chooses a contribution level at one end or the other of the range from 0 to c_j. If player *i*'s parameter of advantageous inequality aversion (β_i) is sufficiently low, then material incentives dominate and player *i* chooses not to contribute to the public good (that is, $c_i = 0$). If player *i*'s parameter of advantageous inequality aversion passes a certain threshold, then concern to avoid favorable inequality dominates and player *i* matches the contribution of player *j* (that is, $c_i = c_j$).

The critical threshold that determines whether player i chooses $c_i = 0$ or $c_i = c_j$ is given by:

$$\beta_i = 1 - m \qquad 3.5$$

This is because $1 - m$ is the material cost to player i of one more unit of contribution, whereas player *i*'s parameter of advantageous inequality aversion (β_i) is the impact of one more unit of contribution on player *i*'s disutility from favorable inequality. When player i is contributing less than player j, then increasing her contribution to the public good by one unit more lowers the difference between her and her coplayer's contributions by 1 (that is, $x_i - x_j$ is reduced by 1) and therefore her disutility from advantageous inequality aversion is reduced by β_i.

Therefore, the Fehr and Schmidt (1999) model makes a very sharp prediction in the two-player linear VCM environment. Either $\beta_i > 1 - m$ and it is optimal for player *i* to match player *j*'s contribution exactly or $\beta_i < 1 - m$

and it is optimal for i to free ride completely. If $\beta_i = 1 - m$, then player i is indifferent between all level of contribution equal to or below that of player j.

3.3 Experimental Design

In this study, we collected data from the student body of the University of Nottingham in Nottingham, England. We had 104 subjects who participated in 7 sessions.[10] Recruitment was done using ORSEE (Greiner, 2015); each session was conducted using zTree (Fischbacher, 2007) and lasted approximately 90 minutes. Among other tasks, subjects anonymously completed three two-player games without receiving feedback from any game until the end of the session.[11] The average payment was £11.50 and was composed of a fixed participation fee and the realized outcomes from one of the games. Subjects had an incentive to treat each game as potentially payoff relevant because it was only at the conclusion of the session that the payoff relevant game was determined by random draw. The three games relevant to this study are:

- The Blanco, Engelmann, and Normann (2011) *Modified Dictator Game* which elicits individual parameters of advantageous inequality aversion, β_i. This elicitation is a revealed preference measure of β_i that is unconfounded by strategic and reciprocal considerations.
- A two-player Fischbacher, Gächter, and Fehr (2001) *Public Good Game played with the strategy method* which determines strategy type ∈

[10] Subject pool statistics are listed in the Supplementary Materials, Section 3.6.1.3.1.
[11] As noted in Chapter 1, the revealed preferences of advantageous inequality aversion – the beta data – and the GASP measures here are the same data reported in Chapter 2. We have supplemented that data with the public goods data for the analysis in this chapter.

{perfect conditional cooperator, imperfect conditional cooperator, free rider, other}. These types are defined below.

- A two-player *one-shot direct response Public Good Game with belief elicitation* from which we observe contribution to the public good and elicit beliefs about coplayer's contribution. The results from this game are used as a check of the behavioral validity of Fischbacher, Gächter, and Fehr (2001) *Public Good Game played with the strategy method* along the lines of Fischbacher et al., (2012).

For both of the public good games, the endowment was 10 and marginal per capita return was 0.8. As indicated above, a value of m satisfying $1 > m > 0.5$ is necessary to create a social dilemma in the two-player linear VCM environment. The specific value we employ matches the extensive literature which uses the four-player game with $m = 0.4$. In such games, the marginal social return from the public good is 1.6. We retain this feature while adopting a two-player group size for reasons explained above in Section 3.2.1 and below in Section 3.3.2. Other studies that have used two-player public good games include Blanco et al. (2011) and Rustagi, Engel, & Kosfeld (2010).

Further details of each of these games are described below. Finally, participants completed a variety of questionnaires including standard sociodemographic questions, the Gosling et al. (2003) Ten-Item Personality Inventory (TIPI) measuring the Big Five personality domains, the Guilt and Shame Proneness Scale of Cohen et al. (2011), the Personal Norm of Reciprocity (PNR) Questionnaire developed by Perugini et al. (2003), and the Cognitive Reflection Test of Frederick (2005).

3.3.1 Parameter of Advantageous Inequality Aversion, β_i

The parameter of advantageous inequality aversion, β_i, plays a critical role in determining public good game contribution according to the Fehr and Schmidt (1999) theory of inequality aversion as discussed in Section 3.2.3. Blanco, Engelmann, and Normann (2011) introduced a method of eliciting a subject's β_i parameter using the Modified Dictator Game and we follow their methodology. This method of eliciting a subject's β_i parameter is unconfounded by strategic and reciprocal concerns because the game is a one-shot interaction between two players where one player is passive.

In the Modified Dictator Game, each subject plays the game as a dictator making 21 binary decisions between a selfish payoff option (20 points for self, 0 points for other) and an equal payoff option (x points for self, x points for other) where $x \in \{0, \dots, 20\}$. The first decision considered is between the selfish option and an equal payoff option of 0 points for both self and other, (0, 0). For the remaining decisions, the equal payoff option increases stepwise by 1 point each up to a final decision where the choice is between the selfish option and an equal payoff option of 20 points for both self and other, (20, 20). If this game was selected as payoff relevant by the random mechanism, the software was programmed to randomly pair together two subjects, assign the role of the dictator to one of them, and implement at random one of the dictator's 21 decisions.

The parameter of advantageous inequality aversion, β_i, is determined by analyzing the switching point whereby a subject moves from choosing the

selfish option to choosing the equal payoff option. Let's say that a subject switches from the selfish option (20, 0) to the equal payoff option when the equal payoff option is (x', x'). This switch point implies that the utility this subject receives from the selfish option (20, 0) is greater than the utility she would receive from the previous equal payoff option $(x' - 1, x' - 1)$. In contrast, the utility she receives from the equal payoff option (x', x') is greater than the utility she receives from the selfish option (20, 0). As such, there must be an equal payoff option between x' and $x' - 1$ where the subject is indifferent between the selfish option and the equal payoff option; let's call that point of indifference $\tilde{x}_i$ and let's further arbitrarily approximate $\tilde{x}_i = x' - 0.5$ for simplicity. Recalling Equation 3.1 above, we can solve for each individual's parameter of advantageous inequality aversion as shown in Chapter 2, Section 2.2:

$$\beta_i = 1 - \frac{\tilde{x}_i}{20}. \tag{3.6}$$

Following Blanco, Engelmann, and Normann (2011), we assign $\beta_i = 0$ to subjects who never switch from the selfish option to the equal payoff option and $\beta_i = 1$ to subjects who select the equal payoff option from the start (0, 0) and never select the selfish option. We further have 18 subjects (17.3%) who switch multiple times in the Modified Dictator Game – which is similar to the proportion (15.3%) observed in the Blanco, Engelmann, and Normann (2011) study. It is difficult to know how to interpret or assign a single value for their parameter of advantageous inequality aversion to these subjects who switched multiple times so we again follow the convention of Blanco, Engelmann, and

Normann (2011) and exclude these non-well-behaved subjects from the remaining analysis.

3.3.2 Conditional Cooperation and Free Rider Strategy Types

We determined contribution strategy type by employing the commonly used public good game played with the strategy method first introduced by Fischbacher, Gächter, & Fehr (2001). After introducing the basic public good game experiment as described in Section 3.2.2, we explained the public good game played with the strategy method. In this game, subjects make two types of decisions: an unconditional contribution and a series of conditional contributions on a contribution table. For the unconditional contribution, they simply decide how many of 10 tokens they want to contribute to the public good. The second type of decision involves filling in a contribution table. Here they decide for each possible contribution of their coplayer how many tokens they want to contribute. Therefore, each subject makes 11 separate contribution decisions conditional on the level of contribution of their coplayer. Complete instructions including screenshots from zTree of the two decisions are available in the Supplementary Materials, Section 3.6.1.1.

If this game was selected as payoff relevant by the random mechanism, we programmed the software to pair two randomly selected subjects together. For one of these subjects her contribution would be her unconditional contribution. For the other subject, her contribution would be determined from her contribution table. More specifically, her contribution would be the corresponding conditional amount which she entered on her contribution table at the level of her coplayer's unconditional contribution.

The public good game played with the strategy method has the advantage of detecting all contribution levels a player would choose in response to different contributions by their coplayer, not just the contribution level they make given the particular belief they happen to hold. It therefore gives a much richer picture of a subject's motivations than can be obtained in a direct response one-shot game and it does so without introducing the confounds that would arise in a repeated game. Further, Fischbacher et al. (2012) have demonstrated the behavioral validity of the four-player public good game played with the strategy method.

The main advantage of using a two-player game in our study is that the concept of conditionality is particularly simple because there is only one other player a subject could condition on. Fischbacher et al. (2001) sidesteps the potential problem of multilateral conditioning by requiring players to condition on the average of their coplayers. While this approach has been shown to work well, it is not understood how players infer a profile of contributions from an average. By using a two-player variant, we completely avoid this issue.

In order to determine types, we analyze each subject's contribution decisions in the contribution table. We report four categories: free riders, perfect conditional cooperators, imperfect conditional cooperators, and others. *Free Riders* are subjects who never make positive contributions on their contribution tables irrespective of their coplayer's contributions. Conditional cooperators are split into two types. *Perfect Conditional Cooperators* are subjects who perfectly match their contributions to those of their coplayer. *Imperfect Conditional Cooperators* are those whose contributions – while not perfectly matching those of their coplayer – show an increasing and often monotonic trend which has a

highly significant (at the 1% level) positive Spearman rank correlation coefficient with the contribution of their coplayer. All remaining subjects are classified as *Others*. While the original Fischbacher, Gächter, & Fehr (2001) study separated the 'Hump-shaped contributions' type – later referred to as 'Triangle Contributors' type in Fischbacher, Gächter, & Quercia (2012) and elsewhere – from the 'Other patterns' type, in this study we group these subjects together for simplicity as the separation does not have any impact on the predictions arising from the Fehr and Schmidt (1999) theory of inequity aversion.

In the original Fischbacher et al., (2001) study, 30% of subjects were classified as free riders, 9% as perfect conditional cooperators, 35% as imperfect conditional cooperators, and 26% as others. Other studies including Fischbacher et al., (2012) have found similar proportions of types. We would expect these studies which had four-player games to have a greater proportion of imperfect conditional coordinators than our two-player game due to the problem of multilateral conditioning described above.

3.3.3 One-Shot Direct Response Public Goods Game with Belief Elicitation

Finally, subjects participated in a one-shot direct response public good game with belief elicitation. This public good game had the same payoff function as described in Section 3.2.2 just like the public good game played with the strategy method in Section 3.3.2. In contrast to the public good game played with the strategy method, subjects made only one decision in this public good game: an unconditional contribution; where just as above, they decided how many of 10 tokens they want to contribute to the public good.

Afterwards, they were asked to estimate their coplayer's contribution. If their estimate was exactly right, they would receive a bonus of £1.50. For every point deviation from the correct result, their bonus would decrease by 50 pence meaning that they would not receive any bonus if their estimate deviated by 3 or more points from the correct amount.

If this game was selected as payoff relevant by the random mechanism, the software was programmed to pair together two randomly selected subjects. The sum of their individual unconditional contributions multiplied by the marginal per capita return would be used to determine their individual income from the public good. Their payoff would be the sum of their individual income from the public good and the remainder of their endowment. In addition, they would receive a bonus as described above according to the accuracy of their estimation of their coplayer's contribution.

As both the public good game and the belief elicitation are paid, there is a potential problem of risk-averse subjects hedging against adverse outcomes in the public good game via their stated beliefs. Blanco, Engelmann, Koch, & Normann (2010) examine this potential problem in depth and determine that hedging confounds are not a significant problem unless hedging opportunities are very prominent. We do not believe hedging to be a significant problem in this study as our subjects participated in several games of which only one allowed for the possibility of hedging. The maximum a subject could earn from hedging in this game was just 30% of her original endowment. Finally, this possibility of hedging was further limited as only one of the four games our subjects participated in was payoff relevant as determined by a random draw at the end of session.

3.4 Results

The question we posed at the start of this study is whether inequality aversion explains free riding and conditional cooperation. Section 3.2.1 showed the Fehr and Schmidt (1999) model proposed in their theory of inequity aversion can account for disutility resulting from inequality in payoffs realized in the two-player public good game. Section 3.2.2 showed that inequality in payoffs exists in public good games insomuch as the contributions of the players differ. Combining these two observations together, Section 3.2.3 showed that the theoretical predictions of the Fehr and Schmidt (1999) model are for players to be perfect conditional cooperators or free riders as their individual parameters of advantageous inequality aversion (β_i) are greater than or less than $1 - m$.

It should be acknowledged that these are sharp predictions which set the bar high in terms of evaluating the ability of the Fehr and Schmidt (1999) model of inequity aversion to explain free riding and conditional cooperation. In particular, there could be measurement error of both the parameter of advantageous inequality aversion and the contribution strategy types both of which could negatively impact our assessment of the Fehr and Schmidt (1999) model. As such, there are less stringent criteria by which one could alternatively evaluate the Fehr and Schmidt (1999) model. However, in this study, we have purposely chosen this severe test to show an extreme case.

In Section 3.4.1, we report our preliminary findings: the prevalence of FGF types and the distribution of the parameters of advantageous inequality aversion. In Section 3.4.2, we look to see whether inequality aversion explains free riding and conditional cooperation on the individual level and compare its

predictive success to that of standard economic theory. We conclude in Section 0 by conducting further analysis and exploring other explanations of conditional cooperation.

3.4.1 Preliminary Findings

Recall that the predictions of the Fehr and Schmidt (1999) model are dependent on the threshold whereby concern to avoid advantageous inequality outweighs the material incentives in the public good games and this threshold directly relates to the marginal per capita return from the public good and each subject's individual parameter of advantageous inequality aversion (see Equation 3.5). As the marginal per capita return in our public good games is 0.8, the critical threshold for the parameter of advantageous inequality aversion is 0.2. Figure 3.1 shows the broad distribution of the parameters of advantageous inequality aversion (β_i) in our subject pool around this critical threshold.

We observe subjects on either side of this critical threshold – represented by the yellow line in the figure – above which concern to avoid advantageous inequality outweighs material incentives in our public good games. We observe that there are approximately five times as many *high beta* types (that is, subjects with $\beta_i > 0.2$) in our sample compared to *low beta* types (that is, subjects with $\beta_i < 0.2$). Recall that the theoretical prediction of the Fehr and Schmidt (1999) model is that players who are high beta types will be perfect conditional cooperators whereas low beta types will be free riders. Our observed distribution of the parameters of advantageous inequality aversion (β_i) implies that we ought to see both of these types in our subject pool.

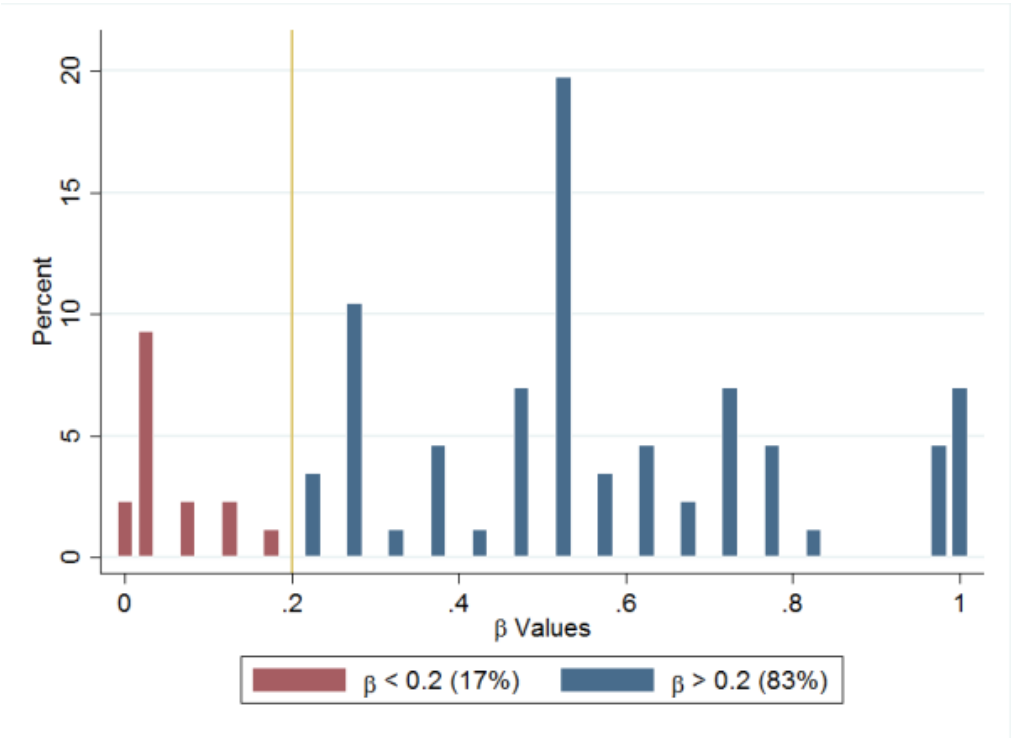


Figure 3.1. Distribution of the parameters of advantageous inequality aversion (β_i). The yellow line indicates the threshold whereby concern to avoid advantageous inequality outweighs material incentives in our public good games given their marginal per capita return of 0.8. In our sample, 17% of subjects are low beta types having $\beta_i < 0.2$ and represented in red. In contrast, 83% of subjects are high beta types having $\beta_i > 0.2$ and are represented in blue.

Figure 3.2 shows the distribution of FGF types and the average contribution by type on the contribution table.[12] The Fehr and Schmidt (1999) model predicted that subjects would either be free riders or perfect conditional cooperators. Here we observe both these theoretically predicted FGF types as well as types not predicted by Fehr and Schmidt (1999) (imperfect conditional cooperators and others). Free riders – subjects who never contribute to the public good irrespective of their coplayer's contributions – account for 27% of our sample. Perfect conditional cooperators who perfectly match their

[12] We include individual contribution schedules according to Beta value and FGF type classification for each subject in the Supplementary Materials, Section 3.6.1.3.3.

coplayer's contributions account for 19% of our sample. The remaining subjects in our sample are classified either as imperfect conditional cooperators – 41% of our sample whose contributions, while not perfectly matching that of their coplayer, show a highly significant (at the 1% level) positive Spearman rank correlation coefficient with the contributions of their coplayer – and others (14%).

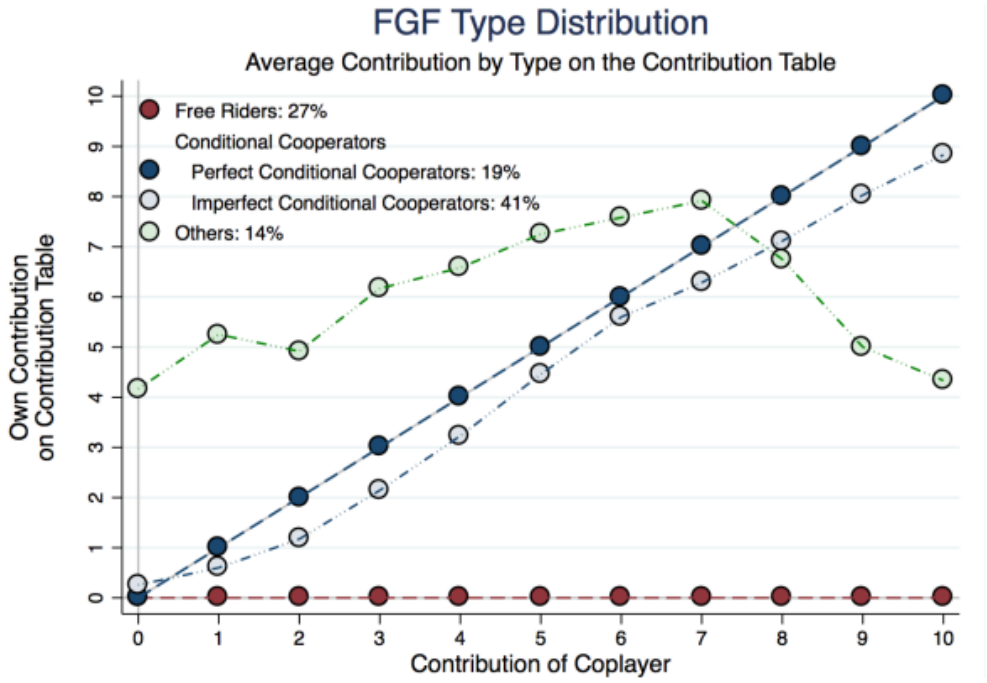


Figure 3.2. FGF Type Distribution. Contribution of coplayer is listed on the horizontal axis while Average Contribution by Type on the Contribution Table is listed on the vertical axis.

Of the 51 conditional cooperators in our subject pool, 16 of them are perfect conditional cooperators (19% of the overall sample and 31% of the conditional cooperator types). The proportion of perfect conditional cooperators is quite high in this study, especially compared with other studies which have investigated the heterogeneity of types such as Fischbacher, Gächter, and Fehr (2001), Cheung (2014), and Hartig, Irlenbusch, & Kölle (2015). As an example,

only 9% of the overall sample and 18% of conditional cooperators were perfect conditional cooperators in Fischbacher, Gächter, and Fehr (2001).

The increased proportion of perfect conditional cooperators compared to previous studies is not necessarily surprising. Recall that previous studies involved four-player public good games in which subjects made contribution decisions based on the average contributions of their three coplayers. If a subject believed that the profile of contributions from the average was made up of heterogeneous contributions and matched her contribution to the lowest one, then such a contribution would necessarily undercut the average. This would result in a greater proportion of imperfect conditional cooperators and a smaller proportion of perfect conditional cooperators. Therefore, it is not surprising that we see fewer imperfect conditional cooperators given that this potential problem of multilateral conditioning is eliminated in the two-player game we used.

To summarize, we observe broad distribution of the parameters of advantageous inequality aversion (β_i) in our subject pool as well as the predicted free rider and perfect conditional cooperator types. However, these two expected FGF types only account for half of our sample and the other half are classified as either imperfect conditional cooperators or other types neither of which are predicted by the Fehr and Schmidt (1999) model.

3.4.2 Does Inequality Aversion explain Free Riding and Conditional Cooperation?

The predictions derived from the Fehr and Schmidt (1999) model are that low beta types will be free riders and high beta types will be perfect conditional cooperators. Our preliminary findings show both a broad distribution of the

parameters of advantageous inequality aversion on either side of the critical threshold as well as the expected free rider and perfect conditional cooperator types among our subjects. However, they also show the imperfect conditional cooperator and other types, contrary to the predictions. Proceeding to our core research question, we see directly the association between types in Table 3.1.[13]

Actual		**Type from FGF**			
		Free Riders	Perfect	*Imperfect*	*Others*
			Conditional Cooperators		
Beta from MDG	$\beta_i < 0.2$	**10** (12%)	**1** (1%)	**3** (4%)	**0** (0%)
	$\beta_i > 0.2$	**12** (15%)	**15** (18%)	**29** (35%)	**12** (15%)

Table 3.1. Association between Types. Each cell lists the frequency of subjects in that category with the overall percentage in brackets. The Fehr and Schmidt (1999) model predicts Low Beta types will be Free Riders (highlighted in red) and High Beta types will be Perfect Conditional Cooperators (highlighted in blue). No other cells are predicted.

Note that only 82 of our 86 subjects with well-behaved preferences are included in this table. The reason is that the Beta values of four subjects were adjacent to the threshold of indifference. Since we arbitrarily selected the midpoint of each possible switching interval in the Modified Dictator Game in

[13] An alternative analysis is found in the Supplementary Materials, Section 3.6.1.3.2. In that analysis we show joint distributions of contributions and β_i values with beliefs of coplayer contribution for all of our subjects.

our calculations of Beta values, we cannot determine a predicted type for these subjects.[14] Given this, we do not include these subjects in Table 3.1.

The rows of Table 3.1 separate subjects into the low and high beta types as determined by revealed preferences in the Modified Dictator Game. The columns separate subjects into types from the FGF Contribution Table responses. The Fehr and Schmidt (1999) model predicts a subset of these possible outcomes: low beta types who are free riders and high beta types who are perfect conditional cooperators. These two cells are highlighted in red and blue, respectively. Free riders with high beta and perfect conditional cooperators with low beta directly contradict the Fehr and Schmidt (1999) model; further, according to the model, neither imperfect conditional cooperator nor other types should be observed. Therefore, the remaining six cells of the table should be empty and as such are shaded gray.

Inequality aversion as modelled by Fehr and Schmidt (1999) explains the free riding and conditional cooperation behavior of 30.5% of our subjects (i.e. 25 out of 82 subjects). Ten of our 82 subjects (12.2%) have low beta and are classified as FGF free riders whereas we would have expected 17% of subjects to be in this category based on the distribution of β_i values in our subject pool. Fifteen of our 82 subjects (18.3%) have high beta and are classified as FGF perfect conditional cooperators whereas we would have expected 83% based on the observed β_i values distribution. Hence, only 30.5% of our subjects are in

[14] Recall that we approximate the point of indifference that defines β_i as the mid-point of a subject's switching interval in the Modified Dictator Game. As we cannot say for sure where in the switching interval a subject's exact point of indifference is, there are a few subjects for whom we cannot rule out that a subject's β_i is equal to the threshold itself or, if not, we are unable to determine on which side of the threshold it lies.

the categories predicted by their parameters of advantageous inequality aversion (β_i) while the majority of our subjects (69.5%) are not.

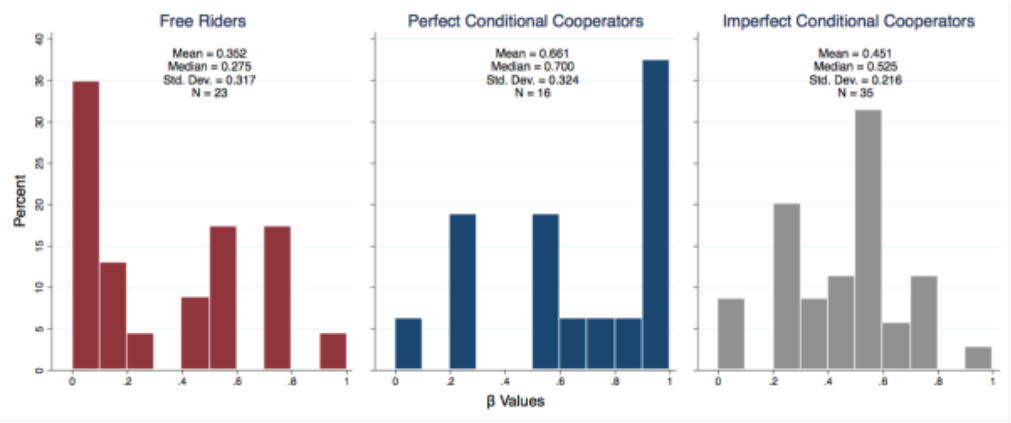


Figure 3.3. Histograms of β Values by FGF Types. The histograms show the distribution of β values by types. Mean, median, standard deviation, and number of observations for each type are also listed. Others are not listed as there are no clear predictions from the Fehr and Schmidt (1999) model of that type, let alone distribution of β values for that type.

We show the distribution of β values for each type with greater specificity in Figure 3.3. Contrary to the predictions of the Fehr and Schmidt (1999) model, we see a wide distribution of β values for the free rider types including many β values greater than 0.2 which are not theoretically predicted. The free rider types in our subject pool are near evenly split between low beta types (45%) as predicted and high beta types (55%) contrary to prediction. It is clear that the Fehr and Schmidt (1999) model fails to explain the observed free riding behavior. Nonetheless, we note that both the mean and median β values of the free rider types are lower compared to those of the perfect and imperfect conditional cooperators.

The distribution of β values for the perfect conditional cooperators is much more in line with the predictions of the Fehr and Schmidt (1999) model. In line with the predictions of the Fehr and Schmidt (1999) model we see 94% of perfect conditional cooperators classified as high beta types (as well as a similar

proportion – 91% – of imperfect conditional cooperators). Overall, the inequality aversion predictions from the Fehr and Schmidt (1999) model do a much better job of explaining perfect conditional cooperation.

Under the strictest definition, inequality aversion accounts for the free riding and conditional cooperation behavior of approximately one third of our subjects. The two main findings contrary to the predictions are the large number of free riders with high beta and the prevalence of types not predicted in the model (imperfect conditional cooperators and others). While accounting for less behavior than expected, it is hard to judge the predictive success of the Fehr and Schmidt (1999) model in isolation. It is better to evaluate the success of inequality aversion in explaining free riding and conditional cooperation relative to that of standard economic theory.

In order to evaluate the relative success of both theories, we use Selten's (1991) measure of predictive success. This measure is calculated by subtracting the area (a measure of precision calculated from the ratio of predicted outcomes compared to all possible outcomes) from the hit rate (the relative frequency of predicted outcomes out of total observations). This measure enables us to compare the relative predictive success of different theories penalizing those which gain success through a loss of precision.

As shown above, the Fehr and Schmidt (1999) model has a hit rate of 30.5%. In contrast, standard economic theory using the selfishness assumption does not recognize inequality aversion. Therefore all subjects are assumed to have $\beta = 0$. Recall that when $\beta_i < 0.2$, it is optimal for a player to free ride completely in our public good game played with strategy method. As such, the

prediction of the standard economic theory using the selfishness assumption is for all subjects to be in top-left cell of Table 3.1. No other cells are predicted. Therefore, the hit rate of standard economic theory using the selfishness assumption is 12.2% (10 out of our 82 subjects). Directly comparing the hit rates of the two models together, we find that the hit rate of the Fehr and Schmidt (1999) model is 18.3 percentage points greater than that of standard economic theory using the selfishness assumption.

However, this comparison does not take into account the area or precision of the two respective models as required by Selten's (1991) measure of predictive success. We calculate the area or precision of the Fehr and Schmidt (1999) model to be 25.0% because the predictions of the Fehr and Schmidt (1999) model fall in just two out of eight possible outcomes.[15] In contrast, the area of standard economic theory is 12.5% (one predicted out of eight possible outcomes). The difference in areas or precision between the two models is therefore 12.5%. Penalizing the hit rate of the Fehr and Schmidt (1999) model for its 12.5% loss of precision, we find that the Fehr and Schmidt (1999) model still has 5.8% more predictive success than standard economic theory using the selfishness assumption. In this sense, the Fehr and Schmidt (1999) model has greater predictive success than standard theory.

[15] We recognize that calculating areas as proportions of cells in Table 3.1 is only one of many possible ways of calculating areas. The conclusions in this section are sensitive to the use of cell counting in Table 3.1 to measure area. Different methods of calculating areas could result in different conclusions.

3.4.3 Further Analyses and Other Explanations

Inequality aversion as modelled by Fehr and Schmidt (1999) offers only limited explanation of free riding and conditional cooperation. While inequality aversion does explain conditional cooperation better than standard economic theory, its overall measure of predictive success shows that there is much room for improvement. In this section we address two possible reasons why inequality aversion might not be more successful. First, it is possible that Beta does not predict type with greater accuracy because the measured strategies of FGF types are not entirely capturing our subjects' true motivations. Therefore, we check the behavioral validity of the two-player public good game played with the strategy method in the manner introduced by Fischbacher et al. (2012). Second, we examine some of the commonly suggested other possible psychological explanations of conditional cooperation alongside inequality aversion and see what new insights we can draw.

One possible explanation for why inequality aversion is not more successful at predicting free riding and conditional cooperation could be that we have incorrectly classified subjects when assigning FGF types. This would be the case if the preferences which determine FGF types do not, or only partially, capture our subjects' true motivations. We can control for this by testing the behavioral validity of the FGF types following the method introduced by Fischbacher et al. (2012).

Recall that our subjects participated in two public good games. In the first public good game – the public good game played with the strategy method – we measure each subject's contribution preferences controlling for their beliefs and

we use these preferences to classify our subjects as free riders, perfect conditional cooperators, imperfect conditional cooperators, and others. The second public good game – the one-shot direct response public good game – is a more standard game where subjects take just one unconditional contribution decision that is based on the beliefs they hold regarding their coplayers contribution. Following their unconditional contribution, we elicit their belief of their coplayer's contribution.

We have thus far assumed, but not shown, that the types we determine from the first game are representative of our subjects' true preferences. To confirm this, we use the belief of coplayer's contribution elicited in the second game to look up our subjects' corresponding conditional contribution in the first game. If the unconditional contribution in the one-shot direct response public goods game matches this conditional contribution in the public good game played with the strategy method (or if there is a difference of at most ± 1; that is, 10% of the total endowment), then this demonstrates consistency between the two public good games. As such, we can assert the behavioral validity of the FGF type measurement with confidence.

Figure 3.4 is a histogram that illustrates this analysis by showing the differences between unconditional and conditional contributions in our two public good games. For each possible deviation, the FGF types are shown in order from left to right starting with free rider types represented by the lightest bars to the left, the perfect conditional cooperators types represented by the darker bars in the middle, and then the imperfect conditional cooperators types represented by light bars to the right (other types are not shown).

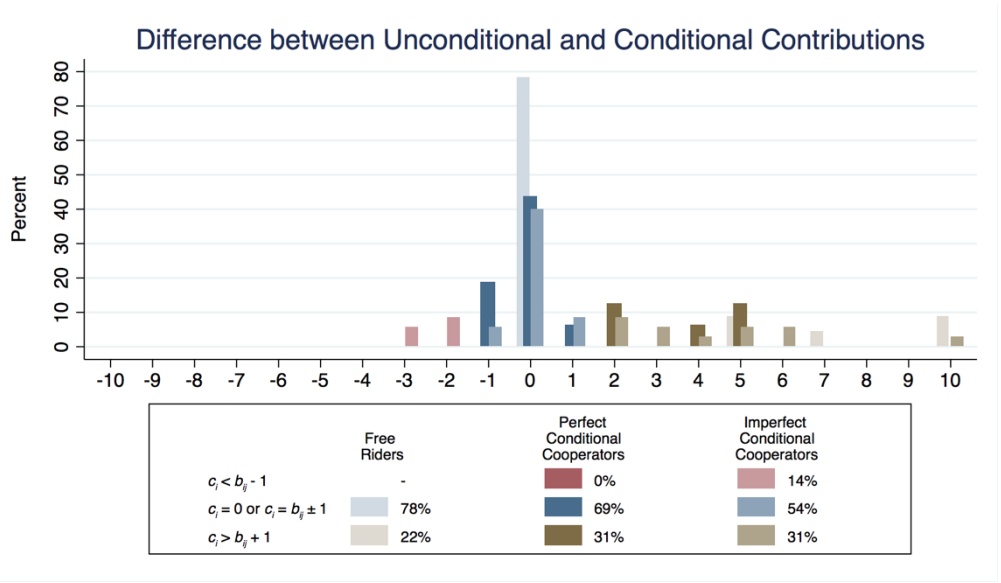


Figure 3.4. Histogram of Difference between Unconditional and Conditional Contributions. In each grouping of bars, the left lightest bars represent free riders, the center darker bars represent perfect conditional cooperators, and the right light bars represent imperfect conditional cooperators. The blue bars represent instances when there is no difference between unconditional and conditional contributions (or when the difference is at most ± 1). Red (brown) bars represent instances when contributions are lower (higher) than predicted. c_i refers to the unconditional contribution in the one-shot direct response public good game of player i. b_{ij} refers to conditional contribution in the public good game played with the strategy method at the same level of belief elicited in the one-shot direct response public good game. We assert the behavioral validity of the FGF type measures to the extent that these two contributions are consistent.

We observe that for nearly 70% of perfect conditional cooperators there is no difference or at most a difference of ± 1 – that is, a deviation of only 10% of the total endowment – between their unconditional contribution in the one-shot direct response public good game and their conditional contribution at the level they believed their coplayer would contribute in the public good game played with the strategy method. We find the same for over half of imperfect conditional cooperators. Further, for free riders – who by definition entered conditional contributions of 0 for every possible coplayer contribution – we find that nearly 80% do not contribute in the one-shot direct response public good game. In comparison, 67% of perfect conditional cooperators, 67% of imperfect

conditional cooperators, and 95% of free riders were consistent between the two public good games in the Fischbacher et al. (2012) subject pool.[16]

When we further analyze deviations greater than 1, we find that perfect conditional cooperators whose direct response and predicted contributions differ by more than 10% always contribute more in direct response than their belief. Contributing more than your belief is directly contrary to the Fehr and Schmidt (1999) model. That is perfect conditional cooperators favor disadvantageous as opposed to advantageous inequality when deviating in the second game.

Likewise, two thirds of deviations for imperfect conditional cooperators are in the same direction indicating a preference for disadvantageous over advantageous inequality. By default, all deviations by free riders must be in this direction and so we only comment that four out of the five FGF free riders who make positive contribution in the second game have $\beta_i > 0.2$ and are therefore classified as high beta. Overall, we conclude that our preferences based measure of FGF types has behavioral validity.

A second possible explanation for why inequality aversion is not more successful at predicting free riding and conditional cooperation is that conditional cooperation might have multiple explanations. If this were the case, then our finding that inequality aversion alone only accounts for the behavior one third of our subjects would be less surprising. So far we have focused on

[16] Fischbacher et al. (2012) only report the percent of consistent free riders in their study. However, they kindly provided us with their original data upon our request and we were able to calculate the consistency of perfect and imperfect conditional cooperators the results of which we report here.

one possible explanation – inequality aversion – in depth but others include psychological personality traits, guilt and shame aversion, reciprocity, and cognitive reflection. As part of the questionnaire we administered at the end of the study, we included measures for some of these additional explanations.[17]

We then conducted a probit analysis of perfect conditional cooperation on these psychological measures as well as standard socio-demographic variables to further examine the causes of conditional cooperation. We analyzed models which include all four sets of psychological measures we thought, a priori, might explain conditional cooperation in the Supplementary Materials. We further estimated separate models removing each set of psychological measure and the group of socio-demographic variables one-by-one (results of which are available in the Supplementary Materials, Section 3.6.1.3.4). We used likelihood ratio tests to compare the goodness of fit of each to the all-inclusive model and find that including the Gosling et al. (2003) Ten-Item Personality Inventory (TIPI) measuring the Big Five personality domains, the Perugini et al. (2003) Personal Norm of Reciprocity Questionnaire measures, or the socio-demographic variables in the original model does not offer a statistically significant improvement to model fit. In contrast, we find that the original model has statistically significant better fit than models where the Cohen et al. (2011) Guilt and Shame Proneness Scale or the Frederick (2005) Cognitive Reflection Test measures are removed.[18]

---

[17] These psychological survey instruments are listed in Supplementary Materials, Section 3.6.1.2.

[18] The better fit of the all-inclusive model is significant at the 1% level when comparing it to models where we remove the Cohen et al. (2011) Guilt and Shame Proneness Scale measures ($\chi^2(2) = 12.77, p = 0.002$) or the Frederick (2005) Cognitive Reflection Test measures

Perfect Conditional Cooperation		
	Coefficients	Marginal Effects
Blanco et al. (2011)'s Measure of Inequality Aversion		
Beta	1.765***	0.367***
	(0.696)	(0.139)
Cohen et al. (2011)'s Guilt and Shame Proneness Scale		
Guilt – Negative Behavior Evaluation	0.230	0.048
	(0.149)	(0.031)
Shame – Withdraw	0.300	0.062*
	(0.185)	(0.038)
Frederick (2005)'s Cognitive Reflection Test		
Cognitive Reflection	0.333*	0.069*
	(0.185)	(0.036)
Constant	-4.369***	
	(1.071)	
N	86	86
χ^2 (4)	16.9***	
Pseudo R^2	0.205	
Log Likelihood	-32.87	

Table 3.2. Probit Analysis of Perfect Conditional Cooperators. Dependent variable is 1 if subject is classified as a perfect conditional cooperator. Variables in Cohen et al. (2011)'s Guilt and Shame Proneness Scale range in value from 1 (low) to 7 (high). Variables in Frederick (2005)'s Cognitive Reflection Test are between 0 and 3 according to number of correct answers. Marginal effects are reported at the means. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Our preferred model is one which includes our measure of inequality aversion as well as the Cohen et al. (2011) Guilt and Shame Proneness Scale and the Frederick (2005) Cognitive Reflection Test measures. Table 3.2 records the results of this analysis. The dependent variable in this analysis is a dummy variable where 1 indicates subjects classified as perfect conditional cooperators

($\chi^2(1) = 7.95, p = 0.005$). The better fit of the all-inclusive model is not statistically significant when compared to models where we remove the Gosling et al. (2003) Ten-Item Personality Inventory (TIPI) measuring the Big Five personality domains ($\chi^2(5) = 9.18, p = 0.102$), the Perugini et al. (2003) Personal Norm of Reciprocity Questionnaire measures ($\chi^2(3) = 6.21, p = 0.102$), or the socio-demographic variables ($\chi^2(3) = 1.85, p = 0.603$).

according to their contribution tables in the public good game played with the strategy method and 0 indicates all other subjects.

In every model, Beta was a positive and highly significant explanatory variable. The greater a subject's aversion to advantageous inequality – as measured by her Beta value – the greater the likelihood she is a Perfect Conditional Cooperator. This result provides further support to the main finding of this study which is that inequality aversion does explain, at least some part of, conditional cooperation.

In addition to finding that a subject's Beta value is positively and highly significantly associated with increased likelihood of being a perfect conditional cooperator, we examined the impact of both psychological measures on conditional cooperation. We find that both greater numbers of correctly answered questions on the Frederick (2005) Cognitive Reflection Test or higher self-assessment on Cohen et al. (2011)'s Guilt and Shame Proneness Scale (that is, greater reported guilt and shame) were associated with increased likelihood of being a perfect conditional cooperator. While neither the guilt nor the shame coefficients were significant in the probit analysis, we find that they are jointly significant in log likelihood tests and that the marginal effect of shame is significant. This analysis suggests that multiple explanations may be working in combination to account for conditional cooperation and suggests potential directions for future research.

3.5 Conclusions

In conclusion, we have shown that inequality aversion as modelled by Fehr and Schmidt (1999) offers only limited explanation of free riding and conditional

cooperation. We find that, for the most part, inequality aversion explains perfect conditional cooperation as predicted by the Fehr and Schmidt (1999) model. In contrast, we find that subjects classified as free riders are equally likely to have low beta (as predicted by the Fehr and Schmidt (1999) model) as they are to have high beta (contrary to prediction). Further, the majority of our subjects are classified as imperfect conditional cooperators or other types; that is, the majority of our subjects are classified as types not at all predicted by the Fehr and Schmidt (1999) model.

While having greater predictive success at explaining these phenomena than standard economic theory, there remains much which inequality aversion is unable to explain. One reason the model has only limited success is that the predictions from inequality aversion as modelled by Fehr and Schmidt (1999) of conditional cooperation and free riding are particularly sharp. In the two-player public good game, subjects who are sufficiently averse to advantageous inequality are predicted to perfectly match their coplayer's contributions whereas those who are not sufficiently averse are predicted to free ride entirely. As such, inequality aversion as modelled by Fehr and Schmidt (1999) is unable to explain the most commonly observed form of conditional cooperation: imperfect conditional cooperation.

If we were to loosen the predicted form of cooperation to include imperfect, as well as perfect, conditional cooperation, the overall success the predictions would significantly improve. Are there theoretical justifications for loosening the definition in this way? We argued above that part of the elegance of the Fehr and Schmidt (1999) model is its simplicity. However, a disadvantage of its simplicity is that its predictions are non-compromising. In contrast, a non-linear

formulation of inequality aversion might allow some forms of imperfect conditional cooperation and thus explain some of the imperfect conditional cooperators we observed. This is an area ripe for further exploration.

In games with more than two players, it is more complicated to derive predictions from the Fehr and Schmidt (1999) model for the public good game with strategy method because of issues such as multilateral conditioning, the possibility of a player experiencing more than one type of inequality at a time, and so on. As such, we focus on the simpler case of two-player public good games in this study. However, we recognize that behavior in groups is different than that between pairs of individuals and therefore we recommend caution in applying our findings to broader cases. In this study, players participate in a game where they can avoid inequality for sure. However, the relevance of inequality aversion in public good games with more than two players is much less straightforward. We expect that the direction of our findings would be similar, but that their magnitude would be less as the issues discussed above would make a player's inequality aversion calculations more complicated. This would be another area rich for future research.

Further, we have shown in our analysis that, in addition to inequality aversion, other psychological measures also have an important role to play in explaining conditional cooperation. While inequality aversion may not be the sole behavioral mechanism by which conditional cooperation functions, we have shown that it has a place of prominence among multiple explanations which working in tandem are more fully able to explain conditional cooperation. We hope to have shown the importance of inequality aversion to the phenomenon of conditional cooperation and to motivate future research

exploring the ways in which inequality aversion interacting with other psychological measures might more fully explain the phenomenon of conditional cooperation.

3.6 Supplementary Materials

3.6.1.1 Instructions

3.6.1.1.1 Background

As described in Chapter 1, the data analyzed in this chapter was collected as part of a larger research project. We analyzed part of the data from the larger project in Chapter 2 and we analyze another portion here. These supplementary materials deal with materials relevant to the current chapter.

3.6.1.1.2 Full Instructions of Relevant Sections

Economic Research Project

You are now taking part in an experiment. If you read the following instructions carefully, you can, depending on your and other participants' decisions, earn a considerable amount of money. It is therefore important that you take your time to understand the instructions. Please do not communicate with the other participants during the experiment. Should you have any questions, please ask us.

The experiment consists of four different sections. In each section you will be called to make one or more decisions. You will have to make your decisions without knowing other participants' decisions in the previous sections. Note further that the other participants will not know your decisions either.

Only one of the sections will be taken into account in determining your final payoff. This will be randomly determined as described below. Each section has the same probability of being selected. You should take your time to make your decision. All the information you provide will be treated anonymously.

The section that will be taken into account in determining your final payment will be selected as follows. **Participant Number 2** was randomly chosen at the very beginning of the experiment. This participant will draw a ball from a cloth bag **after** all participants have completed all sections. Each ball in the cloth bag has a different colour and each colour corresponds to a different section: yellow, blue, green, and red. The resulting colour and corresponding section will be used to calculate your payment.

The computer will randomly pair you with another participant in the room and will assign the roles. The matching and roles assignment will remain anonymous. You will not know which role you were playing until the end of the game.

Your earnings will be paid to you in cash at the end of the experiment at a rate of 1 point = 50 pence. Earnings will be confidential.

Yellow Section

In this section the situation is as follows:

> *Person A is asked to choose between two possible distributions of money between her and Person B in twenty-one different decision problems. Person B knows that A has been called to make those decisions, and there is nothing he can do but accept them.*
>
> *The roles of Person A and Person B will be randomly determined at the end and will remain anonymous*

Before making your decisions please read carefully the following paragraphs.

The decision problems will be presented in a chart. Each decision problem will look like the following:

Person A's Payoff	Person B's Payoff	Decision	Person A's Payoff	Person B's Payoff
20	0	Left Right	5	5

You will have to decide as Person A; hence if in this particular decision problem you choose left, you decide to keep the 20 points for yourself so Person B's payoff will be 0 points. Similarly, if you choose Right, you and Person B will earn 5 points each.

You will need to choose one distribution (Left or Right) in each of the twenty-one rows you will have in the screen. If this is chosen as the payoff relevant section, the computer will randomly choose one of the twenty-one decisions. The outcome in the chosen decision will then determine your earnings.

The computer will randomly pair you with another participant in the room and will assign the roles. The matching and roles assignment will remain anonymous.

Please note that you will make all decisions as Person A but the computer might assign you Person B's role.

If you are assigned the role of A, you will earn the amount that you have chosen for Person A in the relevant situation and the person paired with you will earn the amount that you have chosen for Person B.

In the case that you are assigned the role of Person B, you will earn the amount that Person A whom you are paired with has chosen for Person B in the relevant situation.

Blue Section

[This section was unrelated to this paper.]

Final Sections (Green and Red)

You are now taking part in a new experiment. These instructions are solely for your private information. **You are not allowed to communicate during the experiment.** If you have any questions, please raise your hand. A member of the experimental team will come to you and answer them in private.

All participants will be divided into groups of two members. **Only the experimenters will know who is in which group.**

The Decision Situation

We first introduce you to the basic decision situation. Then, you will complete a pre-study questionnaire on the screen in front of you, which is intended to help you understand the decision situation.

In each group, every member has to decide the allocation of 10 tokens. You can put these 10 tokens into your **private account** or you can put some or all of them into a **project**.

Your income from the private account

You will earn 1 point for each token you put into your private account. For example, if you put all 10 tokens into your private account, your income from your private account would be 10 points. If you put 6 tokens into your private account, your income from this account would be 6 points. **No one except you earns anything from tokens you put in your private account.**

Your income from the project

Each group member will profit equally from the amount you or any other group member put into the project. The income for each group member from the **project** will be determined as follows:

Income from the project *= 0.8 × sum of all contributions*

If, for example, the sum of all contributions to the project by you and the other group member is 15 tokens, then you and the other member of your group each would earn 15 × 0.8 = 12 points out of the project. If both members of the group contribute a total of 10 tokens to the project, you and the other member of your group would each earn 10 × 0.8 = 8 points.

Total income

Your total income is the sum of your income from your private account and from the project:

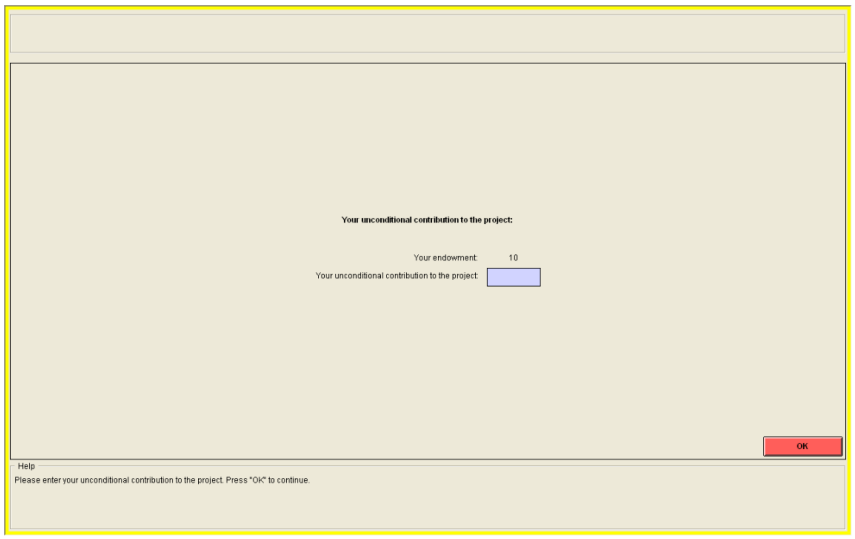
Your total income *= Income from your private account + Income from the project* *=10 – your contribution to the project + 0.8 × sum of all contributions to the project*

Green Section

The experiment is based on the decision situation just described to you, conducted **once.** You will enter your decisions on the screen in front of you.

As you know, you will have 10 tokens at your disposal. You can put them into a private account or into a project. Each subject has to make **two types** of decisions in this experiment, which we will refer to below as the **"unconditional contribution"** and the **"contribution table"**.

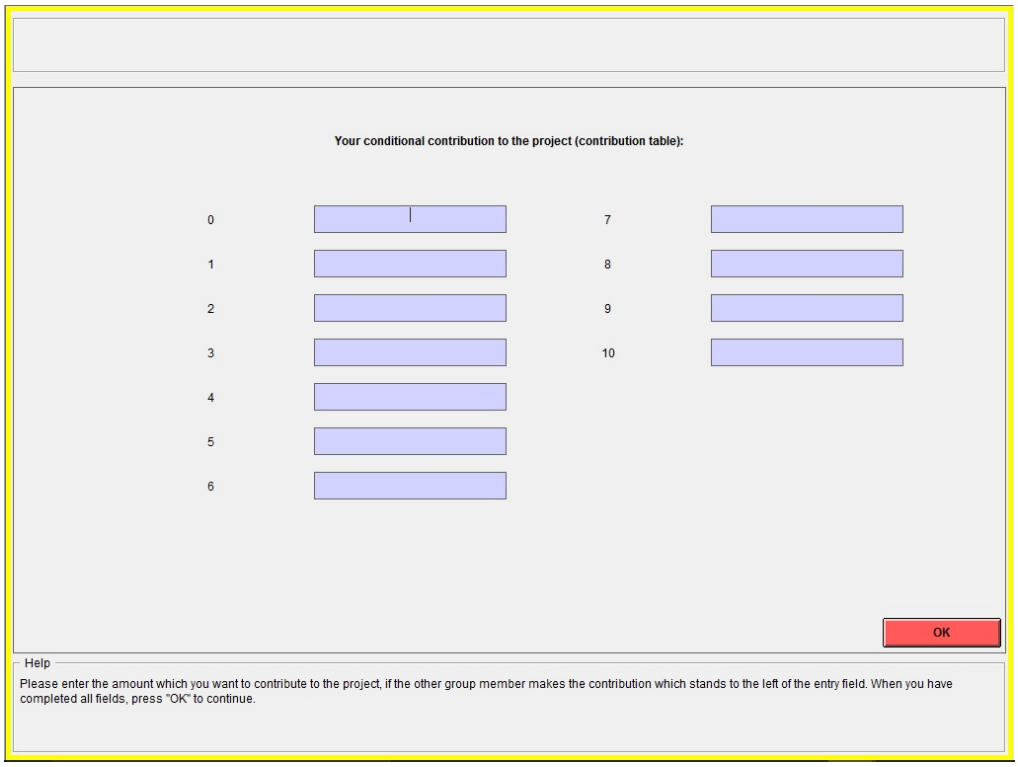
- In the **unconditional** contribution you simply decide how many of the 10 tokens you want to put in the **project**. Please indicate your contribution in the following screen:



After you have determined your unconditional contribution, please click "OK".

- Your second task is to fill in a "contribution table" where you indicate how many tokens **you want to contribute** to the project **for each possible contribution of the other group member**. Here, you can condition your contribution on that of the other group member. This will be immediately clear to you if you take a look at the following table.

This table will be presented to you in the experiment:



The numbers to the left of the blue cells are the possible contributions of the **other** group member to the project. You have to insert how many tokens you want to contribute to the project into each input box – conditional on the indicated contribution by the other member of your group. **You must enter a number between 0 and 10 inclusive in each input box**. For example, you have to indicate how much you contribute to the project if the other group member contributes 0 tokens to the project; how much you contribute if the other group member contributes 1, 2, or 3 tokens; etc. Once you have made an entry in each input box, click “OK”.

After all participants of the experiment have made an unconditional contribution and have filled in his contribution table, a random mechanism will select one member from every group. For **this** group member, it is his **contribution table**

that will determine his actual contribution; whereas, for the **other** group member, it is his **unconditional contribution** that will determine his actual contributions. You will not know whom the random mechanism will select when you make your unconditional contribution and fill in your contribution table. You must therefore think carefully about both decisions because either could determine your actual contribution. Two examples should make this clear.

EXAMPLE 1: Suppose that **the random mechanism selects you;** and that the other group member made an unconditional contributions of 2 tokens. If you indicated in your contribution table that you will contribute 1 token if the other group member contributes 2 tokens, then the total contribution to the project is given by 2 + 1 = 3 tokens. Each group member would, therefore, earn 0.8 × 3 = 2.4 points from the project plus their respective income from their own private account. If, instead, you indicated in your contribution table that you would contribute 9 tokens if the other group member contributes 2 tokens, then the total contribution of the group to the project would be given by 2 + 9 = 11 tokens. Each group member would earn 0.8 × 11 = 8.8 points from the project plus their respective income from their own private account.

EXAMPLE 2: Suppose **that the random mechanism does not select you;** and that your unconditional contribution is 6 tokens. If the other group member whom the random mechanism did select indicates in her contribution table that she will contribute 1 token if the other group member contributes 6 tokens, then the total contribution of the group to the project is given by 6 + 1 = 7 tokens. Each group member will therefore earn 0.8 × 7 = 5.6 points from the project plus their respective income from their own private account. If, instead, the randomly selected group member indicates in her contribution table that she

contributes 9 if the other group member contributes 6 tokens, then the total contribution of the group to the project is 6 + 9 = 15 tokens. Each group member would therefore earn 0.8 × 15 = 12 points from the project plus their respective income from their own private account.

The random selection of the group member whose contribution table will determine his actual contribution will be made as follows. Recall that each group member is assigned a **Group Member ID** that ends with either an odd or even number. **Participant Number 3** was randomly selected at the very beginning of the experiment. This participant will roll a die **after** all participants have completed all sections. If the resulting number is odd, the contribution tables of participants whose ID's end in odd numbers will determine their contribution and the unconditional contributions of participants whose ID's end in even numbers will determine their contribution. If the resulting number is even, that the unconditional contribution of participants whose ID's end in odd numbers will be relevant and the contribution table of group member's whose ID's end in even numbers will determine their contributions.

Red Section

As in the previous experiments you are in a group composed of 2 people.

The decision situation is the same as the one described on the instruction sheet of the previous experiment. Each member of the group has to decide about the usage of the 10 tokens. You can put these 10 tokens into your private account or you can put them fully or partially into a project. Each token you do not put

into the project is automatically placed into your private account. Your income will be determined in the same way as before.

Reminder:

Your total income *= Income from your private account + Income from the project* *=10 – your contribution to the project + 0.8 × sum of all contributions to the project*

The decision screen looks like this:



1. First you have to decide on your contribution to the project, that is, you have to decide how many of the 10 tokens you want to contribute to the project, and how many tokens you want to put into your private account. The other member of your group has to make the corresponding decision. This is the only contribution decision that you or they make in this experiment. There is no contribution table.

2. Afterwards you have to estimate the contribution to the project of the other group member in this period. You will be paid for the accuracy of your estimate:

- If your estimate is exactly right (that is, if your estimate is **exactly** the same as the actual contribution of the other group member), you will get **1.5 pounds** in addition to your other income from the experiment.
- If your estimate deviates by one point from the correct result, you will get 1 additional pound.
- A deviation by 2 points still earns you 0.5 additional pounds.

If your estimate deviates by 3 or more points from the correct result, you will not receive any additional pounds.

Scenarios and Questionnaire

Scenarios

[This section was unrelated to this paper.]

Questionnaire

While calculating your payoff, we would like to ask you to answer the following questionnaire.

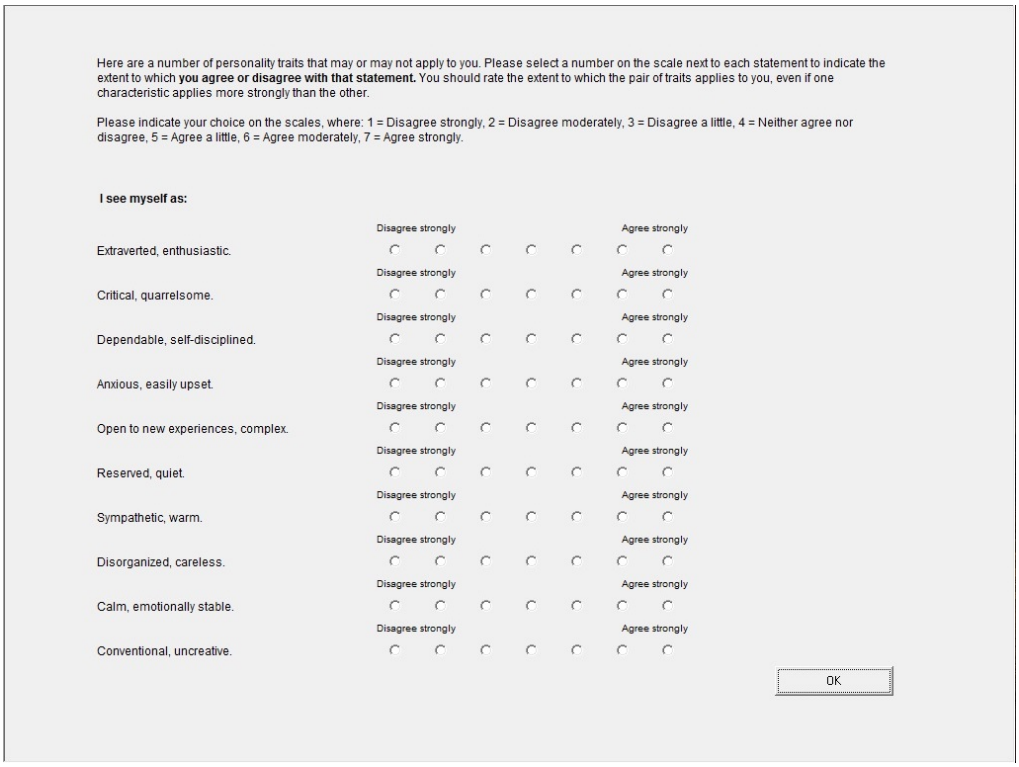
Please answer each of the following questions as accurately as possible. Of course, your answers will be treated confidentially. Your honest answers will

be of immense value for our scientific investigation. Thank you in advance for your cooperation.

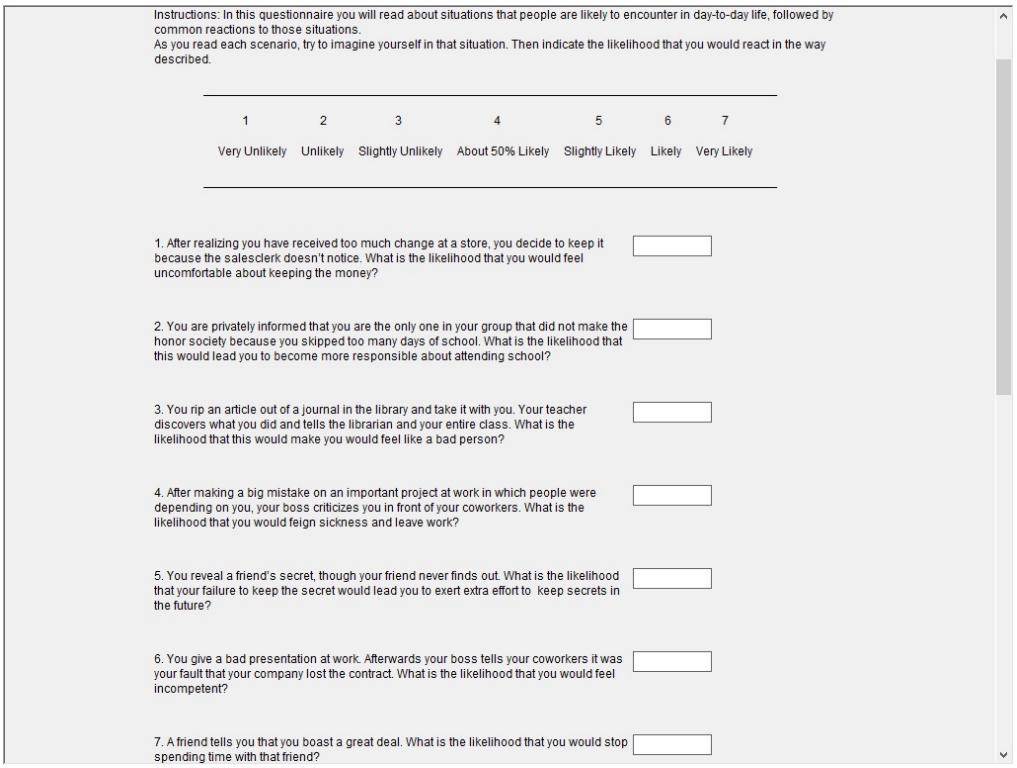
3.6.1.2Psychological Survey Instruments

Below are the psychological survey instruments as presented in zTree to the subjects. For more details on the survey instruments as well as guidance as to how overall scores are determined, please refer to the respective papers.

3.6.1.2.1 Gosling et al. (2003)'s Ten-Item Personality Inventory (TIPI) measuring the Big Five personality domains



3.6.1.2.2 Cohen et al. (2011)'s Guilt and Shame Proneness Scale



Instructions: In this questionnaire you will read about situations that people are likely to encounter in day-to-day life, followed by common reactions to those situations.
As you read each scenario, try to imagine yourself in that situation. Then indicate the likelihood that you would react in the way described.

1	2	3	4	5	6	7
Very Unlikely	Unlikely	Slightly Unlikely	About 50% Likely	Slightly Likely	Likely	Very Likely

1. After realizing you have received too much change at a store, you decide to keep it because the salesclerk doesn't notice. What is the likelihood that you would feel uncomfortable about keeping the money?

2. You are privately informed that you are the only one in your group that did not make the honor society because you skipped too many days of school. What is the likelihood that this would lead you to become more responsible about attending school?

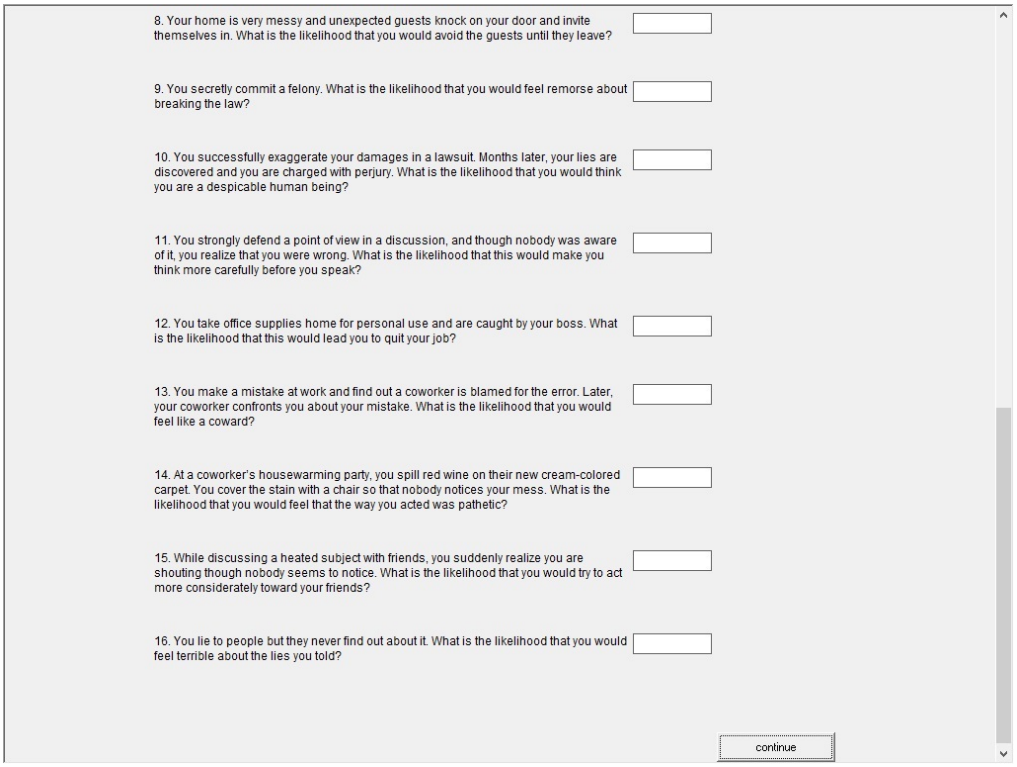
3. You rip an article out of a journal in the library and take it with you. Your teacher discovers what you did and tells the librarian and your entire class. What is the likelihood that this would make you would feel like a bad person?

4. After making a big mistake on an important project at work in which people were depending on you, your boss criticizes you in front of your coworkers. What is the likelihood that you would feign sickness and leave work?

5. You reveal a friend's secret, though your friend never finds out. What is the likelihood that your failure to keep the secret would lead you to exert extra effort to keep secrets in the future?

6. You give a bad presentation at work. Afterwards your boss tells your coworkers it was your fault that your company lost the contract. What is the likelihood that you would feel incompetent?

7. A friend tells you that you boast a great deal. What is the likelihood that you would stop spending time with that friend?



8. Your home is very messy and unexpected guests knock on your door and invite themselves in. What is the likelihood that you would avoid the guests until they leave?

9. You secretly commit a felony. What is the likelihood that you would feel remorse about breaking the law?

10. You successfully exaggerate your damages in a lawsuit. Months later, your lies are discovered and you are charged with perjury. What is the likelihood that you would think you are a despicable human being?

11. You strongly defend a point of view in a discussion, and though nobody was aware of it, you realize that you were wrong. What is the likelihood that this would make you think more carefully before you speak?

12. You take office supplies home for personal use and are caught by your boss. What is the likelihood that this would lead you to quit your job?

13. You make a mistake at work and find out a coworker is blamed for the error. Later, your coworker confronts you about your mistake. What is the likelihood that you would feel like a coward?

14. At a coworker's housewarming party, you spill red wine on their new cream-colored carpet. You cover the stain with a chair so that nobody notices your mess. What is the likelihood that you would feel that the way you acted was pathetic?

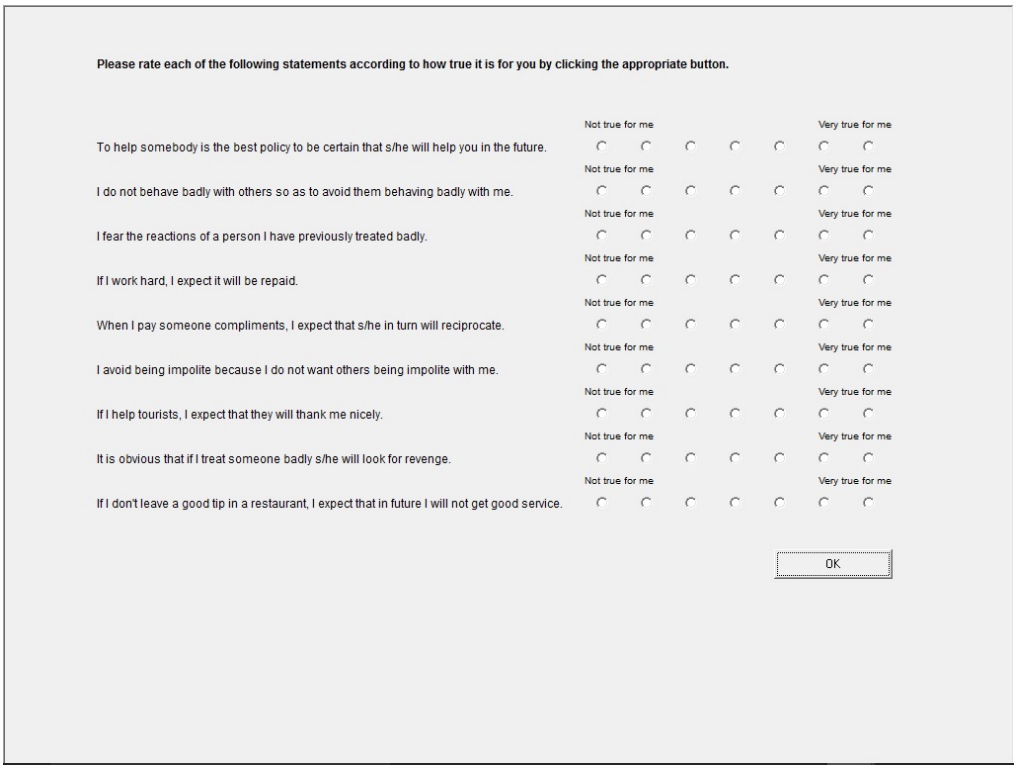
15. While discussing a heated subject with friends, you suddenly realize you are shouting though nobody seems to notice. What is the likelihood that you would try to act more considerately toward your friends?

16. You lie to people but they never find out about it. What is the likelihood that you would feel terrible about the lies you told?

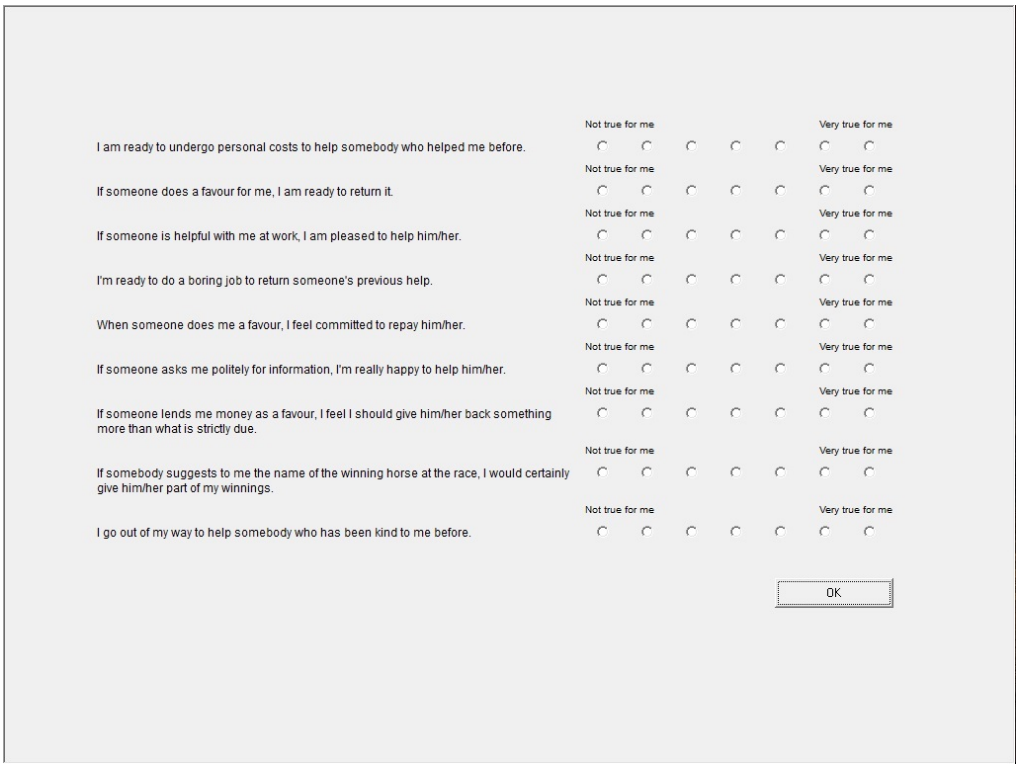
continue

3.6.1.2.3 Perugini et al. (2003)'s Personal Norm of Reciprocity

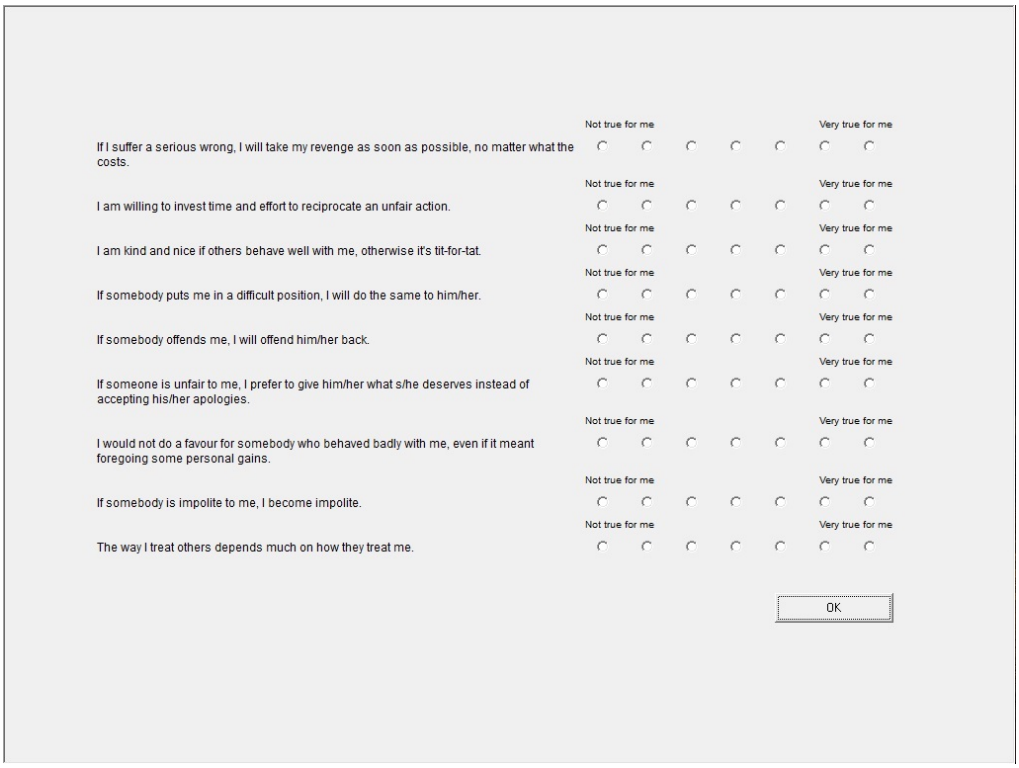
Belief in Reciprocity



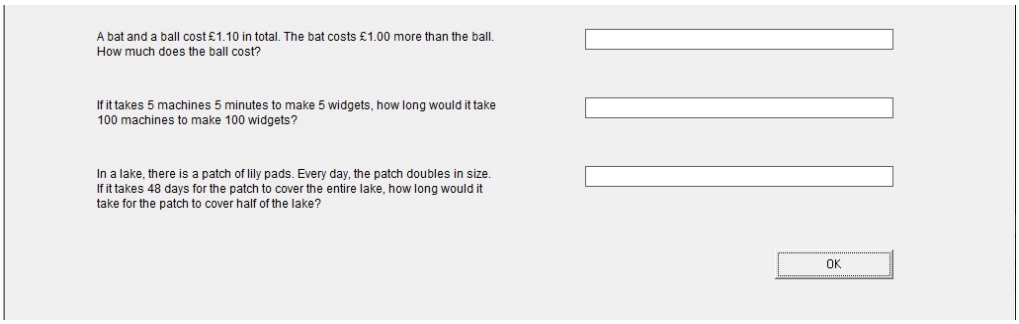
Belief in Positive Reciprocity



Belief in Negative Reciprocity



3.6.1.2.4 Frederick (2005)'s Cognitive Reflection Test



3.6.1.3 Supporting Analysis

We include here supporting analysis of the results from Chapter 3. In Section 3.6.1.3.1 we give subject pool statistics. We analyze the interaction between contributions, Beta values, and beliefs in Section 3.6.1.3.2. In Section 3.6.1.3.3 we show each subject's individual contribution schedule, Beta value, and FGF Type Classification. Finally, we further detail our probit analysis of Perfect Conditional Cooperation in Section 3.6.1.3.4.

3.6.1.3.1 Subject Pool Statistics

In Table 3.3, we list subject pool socio demographic information, Blanco et al. (2011)'s Measures of Inequality Aversion, and Fischbacher, Gächter, & Fehr (2001) Strategy Types. We detail our psychological measures in Table 3.4 including results from Gosling et al. (2003)'s Ten-Item Personality Inventory (TIPI) measuring the Big Five personality domains, Cohen et al. (2011)'s Guilt and Shame Proneness Scale, and Perugini et al. (2003)'s Personal Norm of Reciprocity, and the Frederick (2005)'s Cognitive Reflection Test.

	Overall	Low Beta	High Beta
Social Demographics			
Female	45%	40%	46%
Age	20.20	21.40	19.94
	(1.90)	(2.41)	(1.69)
Minimum	18	19	18
Maximum	29	29	27
Economics	15%	27%	13%
Blanco et al. (2011)'s Measure of Inequality Aversion			
Beta	0.48	0.05	0.58
	(0.29)	(0.05)	(0.23)
Fischbacher, Gächter, & Fehr (2001)'s Strategy Types			
Free Riders	27%	73%	17%
Perfect Conditional Cooperators	19%	7%	21%
Imperfect Conditional Cooperators	41%	20%	45%
Others	14%	0%	17%
Total Sample	86	15	71

Table 3.3. Standard Subject Pool Statistics. Demographic controls, Blanco et al. (2011)'s Measures of Inequality Aversion, and Fischbacher, Gächter, & Fehr (2001)'s Strategy Types. We report means with standard deviations listed in parentheses.

	Overall	Low Beta	High Beta
Gosling et al. (2003)'s Ten-Item Personality Inventory (TIPI)			
Extraversion	4.50	4.10	4.60
	(1.38)	(1.31)	(1.39)
Agreeableness	4.90	5.00	4.90
	(1.00)	(1.15)	(0.98)
Conscientiousness	5.09	4.93	5.12
	(1.26)	(1.25)	(1.27)
Emotional Stability	4.62	4.87	4.57
	(1.42)	(1.19)	(1.46)
Openness to New Experience	5.06	4.47	5.18
	(1.06)	(0.95)	(1.05)
Cohen et al. (2011)'s Guilt and Shame Proneness Scale			
Guilt – Negative Behavior Evaluation	4.73	4.08	4.86
	(1.41)	(1.55)	(1.35)
Guilt - Repair	5.23	4.82	5.32
	(1.15)	(1.41)	(1.08)
Shame – Negative Self Evaluation	5.32	4.98	5.38
	(1.30)	(1.79)	(1.17)
Shame – Withdraw	2.52	2.17	2.60
	(0.97)	(0.72)	(1.01)
Perugini et al. (2003)'s Personal Norm of Reciprocity Questionnaire			
Belief in Reciprocity	4.38	4.27	4.40
	(0.76)	(0.72)	(0.77)
Positive Reciprocity	5.30	4.90	5.39
	(0.81)	(0.77)	(0.80)
Negative Reciprocity	3.33	3.51	3.29
	(1.23)	(1.52)	(1.17)
Frederick (2005)'s Cognitive Reflection Test			
No Questions Right	26%	20%	27%
One Question Right	16%	7%	18%
Two Questions Right	33%	33%	32%
Three Questions Right	26%	40%	23%

Table 3.4. Subject Pool Psychological Measures. Variables in Gosling et al. (2003)'s Ten-Item Personality Inventory (TIPI) measuring the Big Five personality domains, Cohen et al. (2011)'s Guilt and Shame Proneness Scale, and Perugini et al. (2003)'s Personal Norm of Reciprocity Questionnaire range in value from 1 to 7. The Frederick (2005)'s Cognitive Reflection Test is three questions. We report means with standard deviations listed in parentheses.

3.6.1.3.2 Contributions, β_i Values, and Beliefs

An alternative approach to answering our research question is to analyze contributions, Beta values, and beliefs of our subjects. The predictions derived from inequality aversion regarding contributions in the public goods game are that High Beta subjects with $\beta_i > 0.2$ will contribute to the public good

according to their beliefs of their coplayers' contributions whereas Low Beta subjects with $\beta_i < 0.2$ will not contribute regardless of their beliefs.

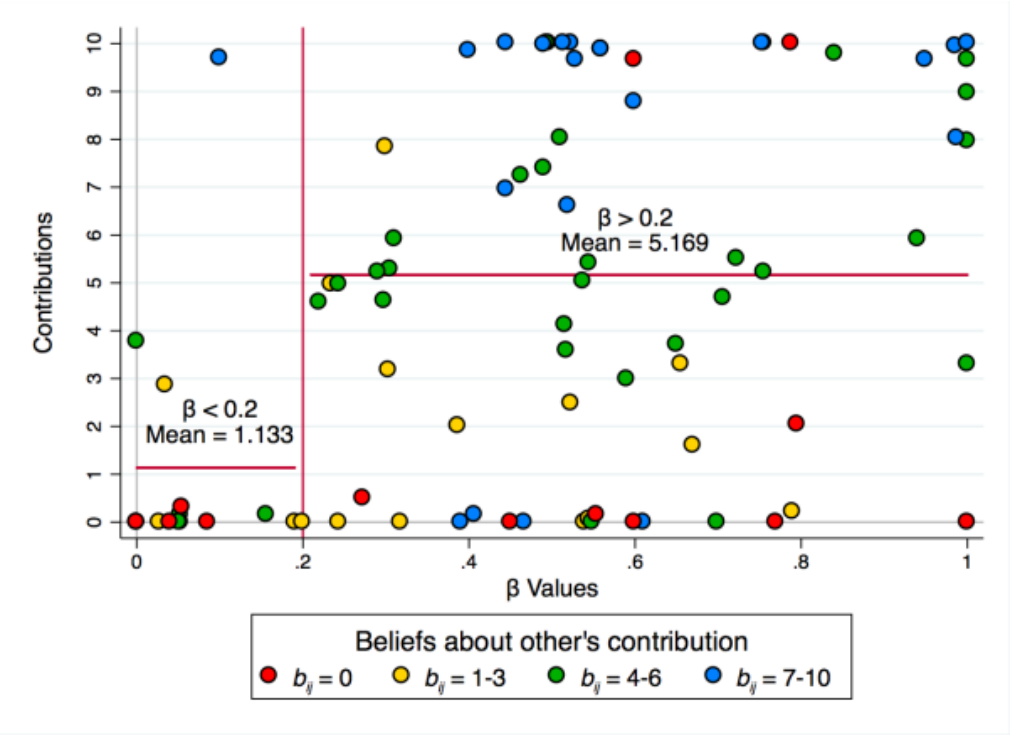


Figure 3.5. Contributions and β_i Values. Joint distribution where each point represents a subject's contribution, c_i, on the y axis and her β_i values on the x axis. The color of point represents each participant's belief of her coplayer's contribution, b_{ij}, with warmer (cooler) colors representing lower (higher) beliefs. The vertical red line separates the *Low Beta* types ($\beta_i < 0.2$) from the *High Beta* types ($\beta_i > 0.2$). The horizontal red lines across the graph are the mean contributions for each type. Note random noise has been added to this distribution to make all the points visible (Stata command: jitter(9)).

Figure 3.5 shows the joint distribution of contributions, c_i, and β_i values for all of our subjects. The color of each point represents the participant's belief of her coplayer's contribution with warmer (cooler) colors representing lower (higher) beliefs. In this analysis, we take the contributions and beliefs from the two-player one-shot direct response Public Good Game with belief elicitation. The vertical red line separates the Low Beta types ($\beta_i < 0.2$) from the High Beta types ($\beta_i > 0.2$). The horizontal red line shows the mean contribution for each β-type. The High Beta types contribute on average 5.2 tokens which as

predicted is more than the Low Beta types. Contrary to prediction, some Low Beta types positively contribute to the public good and the average contribution of their type is 1.133 tokens. The difference in contributions between Low Beta and High Beta types is statistically significant (Kolmogorov-Smirnov test, $d = 0.321, p = 0.009$).

We assume that an inequality averse player would want to match her stated belief about the other player's contribution. However, it is not obvious that this is what an inequality averse player would actually do. Her stated belief about the other player's contribution is likely the mode of a distribution of possible contributions by the other player. Her expected utility maximizing response depends on this full distribution as well as her individual parameters of advantageous and disadvantageous inequality aversion. Recognizing this possibility, we continue the analysis below following our initial assumption.

The majority of the 15 Low Beta subjects do not contribute to the public good as predicted. Only 3 Low Beta subjects made positive contributions to the public good (contributing 3, 4, and 10 tokens, respectively). Two of these made contributions which matched their beliefs of their coplayers' contributions. Only one undercut her coplayer contributing 4 tokens when she believed her coplayer would contribute 5. The average belief of coplayer's contribution for Low Beta types was 2.6 tokens with a standard deviation of 2.9, median belief of 1, and modal belief of 0.

Given that they had passed the critical threshold level of advantageous inequality aversion for possibly contributing to the public good, we expect the High Beta subjects to contribute to the public good such that they match their

beliefs of their coplayers' contributions. The average belief of the High Beta subjects was 4.9 tokens with a standard deviation 3.0 while the median and modal belief was 5 tokens. In fact, the average beliefs of High Beta subjects are close to their average contribution (5.2 tokens) so the prediction seems reasonable in the aggregate.

In Figure 3.5, we see the correlation of individual beliefs of subjects with their contributions. Overall, we see the color of the points shift from warmer to cooler as contributions increase along the y axis. However, we also see some mixture of red points (that is, beliefs of coplayer free riding) at the higher own contribution levels and some blue points (that is, beliefs of high coplayer contributions) at the lower levels of own contribution. The Spearman Correlation coefficient for an individual's contributions and her beliefs of her coplayer's contribution for all the subjects is $\rho = 0.611$, $(p = 0.000)$. Separated by β-type, the Spearman Correlation coefficient for High Beta types is $\rho = 0.550$, $(p = 0.000)$ and for Low Beta types is $\rho = 0.543$, $(p = 0.037)$.

Figure 3.6 is very similar to Figure 3.5 above with the difference being that we have split the mean values into quintiles. We see that the first quintile is significantly lower than the subsequent four quintiles. This result is confirmed by a Kruskall-Wallis test strongly rejects the null hypothesis that the contributions of each quintile are drawn from the same distribution ($\chi^2(4) = 19.78, p = 0.0006$).

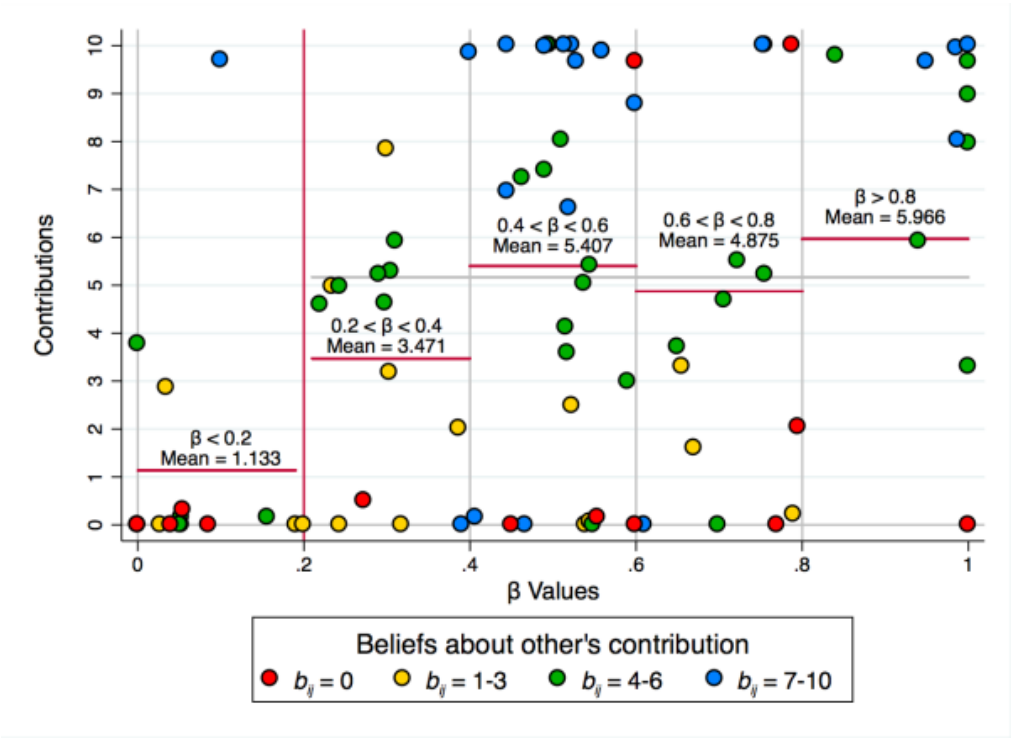


Figure 3.6 Contributions and β_i Values with Mean Values Split into Quintiles. Joint distribution where each point represents a subject's contribution, c_i, on the y axis and her β_i values on the x axis. The color of point represents each participant's belief of her coplayer's contribution, b_{ij}, with warmer (cooler) colors representing lower (higher) beliefs. The vertical red line separates the Low Beta types ($\beta_i < 0.2$) from the High Beta types ($\beta_i > 0.2$). The horizontal red lines across the graph are the mean contributions for each quintile. The grey horizontal line to the right of 0.2 is the mean contribution for all High Beta types. Note random noise has been added to this distribution to make all the points visible (Stata command: jitter(9)).

We show the p-values of the pair wise Wilcoxon rank sum tests comparing the contributions of each quintile in Table 3.5. It is apparent that the distribution of contributions by the Low Beta subjects is statistically significantly different from each of the other quintiles. The distributions of contributions in the other quintiles generally do not often vary one from another in a statistically significant way asides from a few instances in which the distribution of contributions of highest quintile (subjects with $\beta_i \geq 0.8$) varies from a few of the other quintiles.

	$0.2 \geq \beta > 0.4$	$0.4 \geq \beta > 0.6$	$0.6 \geq \beta > 0.8$	$\beta \geq 0.8$
$\beta < 0.2$	0.0128	0.0015	0.0017	0.0003
$0.2 \geq \beta > 0.4$	-	0.1191	0.3764	0.0035
$0.4 \geq \beta > 0.6$	-	-	0.6994	0.1524
$0.6 \geq \beta > 0.8$	-	-	-	0.0863

Table 3.5. *p*-values of pair wise Wilcoxon rank sum tests (two-tailed) comparing contributions between quintiles. Compares the distribution of contributions for each quintile pair wise.

Another possible explanation for the correlation between own contributions and beliefs about the contributions of others is the so-called consensus effect. According to this effect, those who contribute to the public good game have beliefs that are biased to their own type; that is, they are more likely to believe that others will also contribute in the public good game than those who do not contribute. This so-called consensus effect would account for the correlation between own actions and beliefs about those of one's coplayer. Blanco, Engelmann, Koch, & Normann (2014) provide strong evidence of consensus effects in social dilemma games.

3.6.1.3.3 Contribution Schedules per Subject according to Beta value

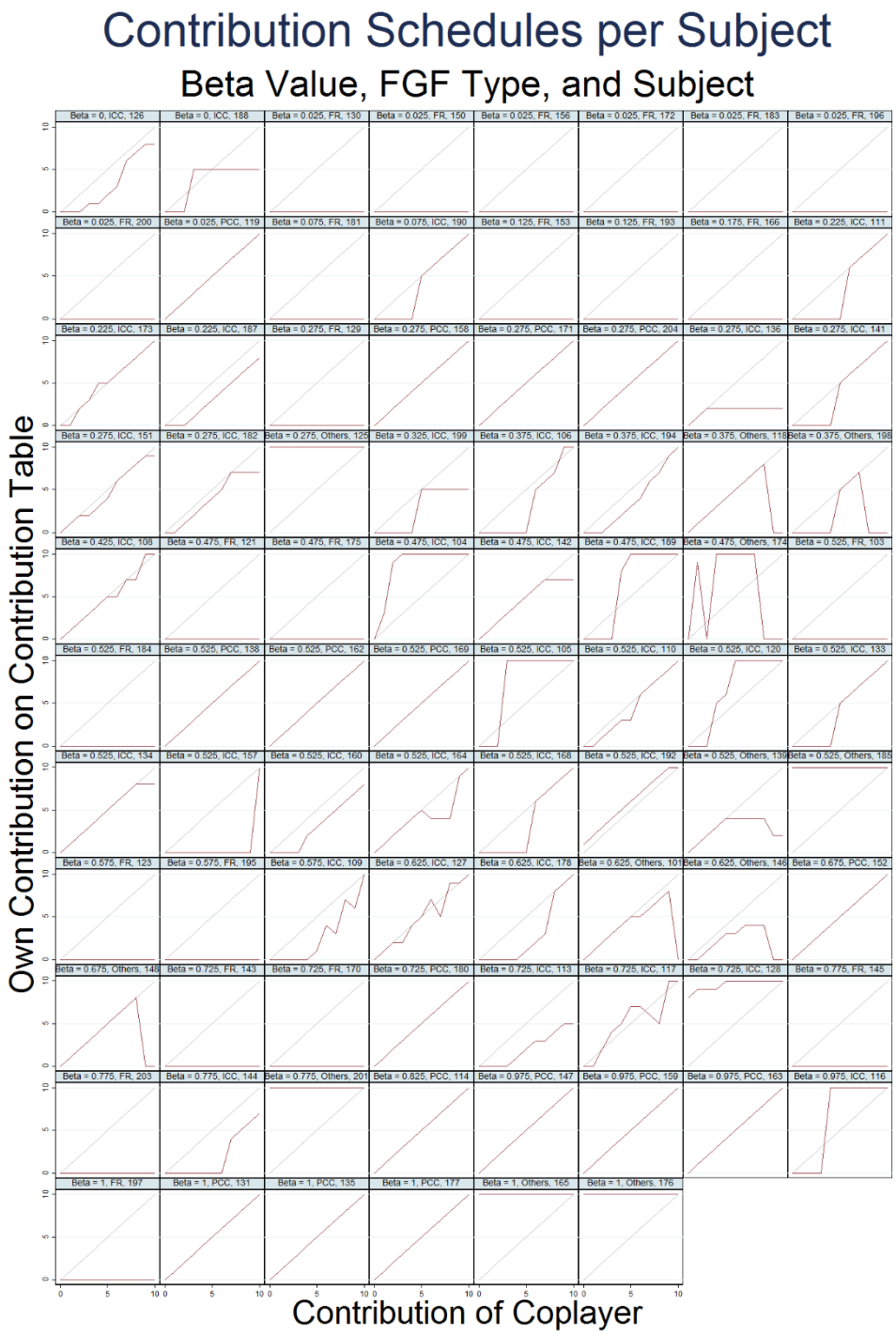


Figure 3.7. Contribution Schedules per Subject. Contribution of coplayer is listed on the horizontal axis while own contribution on the contribution table is on the vertical access and plotted in red. Each plot title includes each subject's individual Beta value, FGF type, and Subject Number. FGF Types include Free Riders (FR), Perfect Conditional Cooperators (PCC), Imperfect Conditional Cooperators (ICC), and Others. The gray 45-degree line shows perfect conditional cooperation.

Figure 3.2 shows average contribution by type on the contribution table for each possible coplayer contribution. While this gives good overview of the contribution schedules of each type, we show each subject's individual contribution schedule in Figure 3.7. Further, we order subjects by their individual Beta values starting from 0 to 1. Recall that Low Beta subjects with $\beta_i < 0.2$ are predicted to be free riders and not contribute to the public good. In contrast, High Beta subjects with $\beta_i > 0.2$ are predicted to be perfect conditional co-operators who match their contributions to those of their coplayers.

3.6.1.3.4 Probit Analysis of Perfect Conditional Cooperators

As described above in Section 0 we conducted a probit analysis to further analyze other possible explanations of conditional cooperation. We include inequality aversion, four psychological measures, and demographic controls in this analysis. The four psychological measures we include are Gosling et al. (2003)'s Ten-Item Personality Inventory (TIPI) measuring the Big Five personality domains, Cohen et al. (2011)'s Guilt and Shame Proneness Scale, the Perugini et al. (2003)'s Personal Norm of Reciprocity Questionnaire, and the Frederick (2005)'s Cognitive Reflection Test. While we had a priori justifications to include each of these measures in our analysis we were not certain whether the inclusion of any of them (or more likely a combination of them) would aid to our understanding of the mechanisms by which conditional cooperation operates and therefore we estimated an initial model including all of the measures and then we estimated a series of models removing each measure one by one. The results of this analysis are listed in Table 3.6. We

Perfect Conditional Cooperators							
Coefficients	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**	**Model 6**	**Model 7**
Blanco et al. (2011)'s Measure of Inequality Aversion							
Beta	3.383***	2.013**	1.995**	2.075***	1.868**	2.876***	1.765**
	(1.185)	(0.816)	(0.832)	(0.790)	(0.850)	(1.048)	(0.696)
Gosling et al. (2003)'s Ten-Item Personality Inventory (TIPI)							
Extraversion	0.453**		0.165	0.214	0.25	0.353*	
	(0.227)		(0.160)	(0.172)	(0.179)	(0.203)	
Agreeableness	0.496*		0.337	0.304	0.317	0.474*	
	(0.295)		(0.228)	(0.241)	(0.261)	(0.275)	
Conscientiousness	0.397*		0.164	0.161	0.125	0.363*	
	(0.221)		(0.183)	(0.185)	(0.173)	(0.206)	
Emotional Stability	-0.0901		-0.106	-0.0424	-0.0606	-0.106	
	(0.149)		(0.134)	(0.141)	(0.139)	(0.146)	
Openness to Experience	0.118		0.1	0.108	0.122	0.116	
	(0.213)		(0.190)	(0.199)	(0.197)	(0.207)	
Cohen et al. (2011)'s Guilt and Shame Proneness Scale							
Guilt – NBE	0.499*	0.243		0.192	0.359	0.475*	0.230
	(0.269)	(0.203)		(0.174)	(0.228)	(0.251)	(0.149)
Shame – Withdraw	0.641**	0.373*		0.461**	0.391*	0.528**	0.300
	(0.258)	(0.210)		(0.216)	(0.214)	(0.238)	(0.185)
Perugini et al. (2003)'s Personal Norm of Reciprocity Questionnaire							
Belief in Reciprocity	-0.695*	-0.263	-0.276		-0.428	-0.528*	
	(0.359)	(0.260)	(0.292)		(0.296)	(0.306)	
Positive Reciprocity	-0.545	-0.137	-0.0973		-0.263	-0.474	
	(0.374)	(0.289)	(0.256)		(0.324)	(0.355)	
Negative Reciprocity	0.305	-0.00862	-0.0375		0.154	0.161	
	(0.287)	(0.222)	(0.202)		(0.250)	(0.248)	
Frederick (2005)'s Cognitive Reflection Test							
Cognitive Reflection	0.687**	0.351*	0.373*	0.454**		0.602**	0.333*
	(0.279)	(0.203)	(0.207)	(0.225)		(0.238)	(0.185)
Demographic Controls							
Age	0.13	0.0494	0.0482	0.0524	0.0624		
	(0.120)	(0.100)	(0.099)	(0.104)	(0.110)		
Female	0.293	0.0646	-0.1	-0.00561	-0.257		
	(0.491)	(0.399)	(0.409)	(0.422)	(0.413)		
Economics	0.282	0.233	0.103	0.287	0.411		
	(0.565)	(0.508)	(0.522)	(0.511)	(0.513)		
Constant	-12.92***	-3.96	-5.056	-9.816***	-6.940*	-9.061***	-4.369***
	(4.803)	(2.808)	(3.229)	(3.462)	(3.675)	(3.349)	(1.071)
N	86	86	86	86	86	86	86
χ^2	28.35**	19.17**	15.58	22.14**	20.39	26.49	16.9
Degrees of Freedom	15	10	13	12	14	12	4
Pseudo R^2	0.343	0.232	0.189	0.268	0.247	0.321	0.205
Log Likelihood	-27.14	-31.73	-33.53	-30.25	-31.12	-28.07	-32.87

Table 3.6. Probit Analysis of Perfect Conditional Cooperators, Coefficients from All Models. Dependent variable is 1 if the subject was classified as a perfect conditional cooperator. Variables in Gosling et al. (2003)'s Ten-Item Personality Inventory (TIPI), Cohen et al. (2011)'s Guilt and Shame Proneness Scale, and Perugini et al. (2003)'s Personal Norm of Reciprocity Questionnaire range in value from 1 to 7. Variables in Frederick (2005)'s Cognitive Reflection Test are between 0 and 3 according to number of correct answers. Age is in years; Female and Economics are dummy variables. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

include seven models along with their respective coefficients and the summary statistics of the regressions in the table.

In order to select the most appropriate model, we performed likelihood-ratio tests comparing each nested model with the original model. As described above, Models 2-7 are all nested in the original model. When the *p*-value of this test is significant, then we interpret the original model which included the eliminated psychological measure as offering a statistically significant improvement in model fit. The results of the likelihood-ratio tests are given in Table 3.7. Models which included measures from the Cohen et al. (2011)'s Guilt and Shame Proneness Scale and the Frederick (2005)'s Cognitive Reflection Test had significantly better fit than models in which those measures were not included. We further estimate Model 7 with just the psychological measures which improved model fit. In this instance, Model 7 which includes only these psychological measures has better fit than the original model with all measures.

Likelihood-ratio tests			
Original Model	*Nested Model*		
Model 1	Model 2	$\chi^2(5) = 9.18$	$p = 0.1021$
Model 1	Model 3	$\chi^2(2) = 12.77$	$p = 0.0017$
Model 1	Model 4	$\chi^2(3) = 6.21$	$p = 0.1019$
Model 1	Model 5	$\chi^2(1) = 7.95$	$p = 0.0048$
Model 1	Model 6	$\chi^2(3) = 1.85$	$p = 0.6032$
Model 1	Model 7	$\chi^2(11) = 11.44$	$p = 0.4069$

Table 3.7. Likelihood-Ratio Tests comparing Models.

The two variables from the Cohen et al. (2011)'s Guilt and Shame Proneness Scale lose their statistical significance in our preferred model so in order to test their joint significance we estimate models which remove this measure and then perform a likelihood-ratio test comparing this nested model

with the original model. We find that the Cohen et al. (2011)'s Guilt and Shame Proneness Scale variables are jointly significant as the original model which includes these psychological variables is preferred to the nested model $(\chi^2(2) = 7.07,\ p = 0.0292)$.

Finally, we recognize the difficulties in interpreting the coefficients from probit estimations. In contrast, the marginal effects are much more understandable. As such, we report the marginal effects at the means for our preferred model in Table 3.8. The significance of the Blanco et al. (2011)'s Measure of Inequality Aversion and the Frederick (2005)'s Cognitive Reflection Test measures do not change when we move from coefficients to marginal effects which gives us greater confidence in the models. Additionally, the marginal effects of the Shame variable from the Cohen et al. (2011)'s Guilt and Shame Proneness Scale are significant. We also briefly introduce and discuss each of these significant measures below.

Marginal Effects	**Model 7**
Beta	0.367***
	(0.139)
Guilt – NBE	0.048
	(0.031)
Shame – Withdraw	0.062*
	(0.038)
Cognitive Reflection	0.069*
	(0.036)
N	86

Table 3.8. Probit Analysis of Perfect Conditional Cooperators, Marginal Effects for Model 7. Dependent variable is 1 if the subject was classified as a perfect conditional cooperator. Variables in Cohen et al. (2011)'s Guilt and Shame Proneness Scale range in value from 1 (low) to 7 (high). Variables in Frederick (2005)'s Cognitive Reflection Test are between 0 and 3 according to number of correct answers. Marginal effects are reported at the means. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Guilt resulting from not cooperating or shame from free riding might be further possible explanations of conditional cooperation. In order to investigate

this possibility, we use Cohen et al. (2011)'s Guilt and Shame Proneness Scale. This measure is made up of two guilt subscales which measure negative behavior evaluation and repair action tendencies and two shame subscales which measure negative self-evaluation and withdrawal tendencies. We report associations between the parameters of inequality aversion and this measure in Chapter 2.

Given that these subscales are known to correlate one with another, we analyzed each individually. Two of the four subscales were significant in these individual analyses: the Guilt – Negative Behavior Evaluation and the Shame – Withdrawal subscales. Recognizing the correlation between these two subscales (Spearman's $\rho = 0.2639,\ p = 0.041$), we include both in Table 3.8 for ease of reporting.

In Cohen et al. (2011)'s Guilt and Shame Proneness Scale subjects imagine themselves in a series of 16 scenarios and indicate how likely they are to respond in the described way. The responses are then averaged according to subscale. The Guilt – Negative Behavior Evaluation subscale measures a moral disposition; that is, it measures the proneness by which a person feels bad for having undertaken certain actions. It is a measure of anticipated or realized guilt. The Shame – Withdrawal subscale measures the proneness by which a person responds to actions self-assessed as shameful by hiding or even removing oneself from public. It describes the tendency towards shame induced avoidance behaviors. Higher scores on the Shame Withdrawal subscale increased the likelihood subjects were perfect conditional cooperators.

Also, we look at the association between reasoning style and conditional cooperation. Stanovich & West (2000) provide a helpful summary of the reasoning literature and describe the two different systems of reasoning in dual-process theories. “System 1” processes are relatively fast, computationally undemanding processes where reasoning is automatic and largely unconscious. In contrast, “System 2” processes are much slower, analytic processes which demand higher cognitive capacity and conscious reasoning.

We use Frederick's (2005) three item Cognitive Reflection Test (CRT) scored 0 to 3 according to the number of correct responses given. Each item in this test has an intuitive, but wrong answer that effortlessly comes to mind. In contrast, the correct answer requires one to ignore this erroneous “System 1” and engage with concentration in the work and effort of “System 2” processes to come up with the true answer. Scores on the cognitive reflection test have been shown to correlate positively and significantly with other measures of cognitive ability or academic achievement.

We find that for each additional item correctly answered on the CRT there is an increase in likelihood that a subject is a perfect conditional cooperator. It’s not clear that there is a “correct” answer to the contribution social dilemma in a public good game, so this is not saying that subjects with higher CRT scores were getting it right while those without were not. However, those with higher CRT scores were more likely to display contribution tables that eliminate inequality than ones that maximize their own payoff.

Chapter 4

Culture of Honor Revisited: Examining Punishment Behavior of US Southerners in Social Dilemmas[19]

4.1 Introduction

There is mounting evidence from tightly controlled experiments that culture and societal background influence social behavior across a variety of settings. Experiments have revealed variation in cooperation in bargaining games across 15 small-scale societies (Henrich et al., 2005), costly punishment across 15 populations (Henrich et al., 2006), antisocial punishment across 16 societies

[19] This chapter draws on joint work in progress with Simon Gächter and Lucas Molleman.

(Herrmann et al., 2008), and intrinsic honesty and the prevalence of rule violation across 23 countries (Gächter & Schulz, 2016).

Variation in cooperativeness across and within societies increases the difficulty individuals face when solving cooperation problems (Fischbacher & Gächter, 2010; Hartig et al., 2015; van den Berg, Molleman, Junikka, Puurtinen, & Weissing, 2015). In many cases, socially optimal solutions to social problems are only achievable when unrelated individuals choose cooperation over their own self-interest. One mechanism which has been shown to increase cooperation in public good games is punishment (Gächter, Herrmann, & Thöni, 2010) and in fact some early research of public good games with punishment showed strong directionality of punishment towards free riders resulting in increased cooperation (Fehr & Gächter, 2000).

The finding that the introduction of punishment to public good games increased cooperation was initially thought to be universal.[20] However, recent studies such as Herrmann et al. (2008) and Fatas & Mateu (2015) showed much greater variation in punishment norms across societies. They found some societies where the punishment of cooperators is as frequent and as intense as the punishment of free riders possibly motivated by revenge and as such the cooperation enhancing effects of punishment were eliminated. The cooperation enhancing effects of punishment seem dependent on subject pool and this begs the question of what exactly is leading to these differences across societies.

[20] These subjects were from a society with a fairly homogeneous culture and norms of punishment. Henrich, Heine, & Norenzayan (2010) describes these subjects as WEIRD – Western, Educated, Industrialized, Rich, and Democratic.

The aim of this paper is to examine whether this type of variation in punishment behavior exists within a given society. Further, we aim to see if one possible explanation for this cross-cultural variation – revenge as motivated by a *Culture of Honor* – might be operating within a given society otherwise known to be highly cooperative, the United States of America. In particular, we investigate whether there is variation in punishment behavior within the United States between Southerners and Non-Southerners. We offer a cleaner test in answering this question than the studies discussed above as our subjects – coming from the same society – have minimal distance aside from our variable of interest. Previous work investigating this phenomenon across societies has additional confounding factors which impact punishment behavior beyond belonging to a culture of honor.

Within-society variation in social behavior has been shown to matter for the outcome of social interactions in studies by Nettle, Colleony, & Cockerill (2011) and Lamba & Mace (2011) who find variation in social preferences within the same city and same society, respectively, according to variations in socioeconomic, demographic, or environmental factors. The advantage of these within-society studies is that many factors which might otherwise vary between subject pools – language, culture, governance, education, etc. – are held constant. In our within-society study, we are able to better isolate our variable of interest – belonging to a *Culture of Honor* – and look more specifically at its impact because of the minimal distance between our comparative subject pools.

4.2 Culture of Honor

Cultures of honor are found in many societies around the world. In these cultures, honor is defined as "the value of a person in his own eyes, but also in the eyes of his society" (Pitt-Rivers, 1965). Honor is of utmost importance and establishes precedence in social standing. Cultures of honor are hypothesized to arise out of harsh socioecological environments where livelihoods rely on resources vulnerable to theft and people are without easy access to legal recourse (Chu, Rivera, & Loftin, 2000; Linquist, 2015). Inhabiting such an environment can lead to the development of a culture where honor is a proxy for the ability and willingness of one to defend oneself, one's family, and one's resources (Pitt-Rivers, 1965).

Honor here is highly interactive and characterized as part of a zero sum, or even "less than zero sum," social game where it can be easily lost but gained back only with great difficulty (Elster, 1990). Honor in this setting is distinct to that associated with social image. According to social image, honor is conferred to the extent that one does good according to social norms while shame is conferred to the extent that one behaves selfishly (Benabou & Tirole, 2006).

This contrasts with the particular psychological phenotype that develops in cultures of honor where any challenge to or impugning of one's honor requires vindication most often through violent aggression. This violent response is necessary to ensure one does not gain a reputation as someone easily taken advantage of. People in cultures of honor are characterized as having a "fight or flight" response to any perceived insult, a willingness to escalate confrontation,

and greater readiness for and willingness to engage with violence. They are characterized by strong and specific norms of revenge (Reed, 1982).

It is not hard to imagine the emergence of a culture of honor among nomadic herders on lawless frontiers. Historically, parts of the frontier in the US South had this very type of socioecological environment which is associated with emergence of the culture of honor. Further, there is an increasing body of evidence that suggests such a culture of honor has been passed from these historic origins and continues to exist even today within modern day America, a society known to be highly cooperative. The existence of a culture of honor in a society can fundamentally affect modern day human behavior, and its social and economic implications are studied in a range of disciplines including social, evolutionary, and cross-cultural psychology; (evolutionary) anthropology; and economics.

Many people who settled in the US South immigrated from northern Ireland and the Scottish Highlands, two areas historically associated with violence and the culture of honor (Grosjean, 2014; Nisbett, 1993). These historic origins have led to many modern day manifestations of the culture of honor in the US South. For example, modern day homicide rates within the US are highest in areas of the South which were associated with these historically harsher socioecological environments (Nisbett, 1993) or which were settled by immigrants from places associated with cultures of honor (Grosjean, 2014). Models show how a culture of honor can be transmitted from these historic origins across generations to its present day manifestations (Linquist, 2015; McElreath, 2003; Nowak, Gelfand, Borkowski, Cohen, & Hernandez, 2015). Social scientists have found that Southerners are more likely than Non-Southerners to accept the use of violence

in sanctioning and socializing children and own guns (Cohen & Nisbett, 1994); be exposed to school violence (Brown, Osterman, & Barnes, 2009); have different cognitive, emotional, behavioral, and physiological responses to insult in social psychology experiments (Cohen, Nisbett, Bowdle, & Schwarz, 1996); and experience greater institutional acceptance of culture of honor motivated violence in field experiments (Cohen & Nisbett, 1997).

While this previous research has built a substantial case for the existence of a culture of honor in the US South, it is heavily reliant on associations found in surveys between attitudes, reported violence, and geography or data observed in experiments run in naturalistic settings, where people are not aware they were in an experiment. Here, we present economic experiments to study behavioral aspects of a culture of honor. This is a promising avenue, given that economic experiments have been successfully used to study culture-specific social norms (Camerer & Fehr, 2004; Fehr & Fischbacher, 2004; Henrich et al., 2005; Gächter et al., 2010). Economic experiments enable us to observe actual behaviors as opposed to reported attitudes.

In this paper, we study the variation between Southerners and Non-Southerners in punishment behavior after a social dilemma. In order to do this, we conduct experiments online using the Amazon Mechanical Turk (MTurk) platform to recruit a large pool of non-students from across the United States.[21] Our subjects take part in an anonymous two-player social dilemma with punishment. In the first stage, participants simultaneously make decisions in a

[21] See Chandler, Mueller, & Paolacci (2014) for a sceptical view and Horton, Rand, & Zeckhauser (2011) for a supportive view on the use of MTurk in experimental economics.

prisoner's dilemma game framed as a stealing game. The socially optimal outcome of the game is for both players to cooperate (i.e. "not take"). However, defection (i.e. "taking") increases a player's individual payoff at the expense of her coplayer. Mutual defection leads to lower payoffs than the socially optimal outcome, but greater individual payoffs than what one would receive if one was defected against while cooperating.

Each player makes their decision in the social dilemma anonymously and after each has decided the results of the first stage are revealed to both. Thereafter, there are two sequential rounds of punishment. In the initial punishment stage, one player is randomly selected as first mover and she has the option to exercise costly punishment against her coplayer regardless of their decision in the social dilemma. Then, in the final stage the other player, after being informed of the first mover's punishment decision, has the option as second mover to exercise costly punishment too. The Nash equilibrium of this game is that both players would choose to defect or "take" and neither would choose to punish.[22]

The frame as a stealing game is designed to reflect the socioecological environment which gives rise to the culture of honor. The taking action (defection) is designed to be a challenge to the honor of subjects from cultures of honor in addition to being a real material loss. Our primary focus of interest is the difference in punishment rates between Southerners and Non-Southerners.

[22] Mengel (2014) provides a helpful overview of the literature of social dilemma games.

In particular, we focus on first mover prosocial punishment – that is, the punishment of those who chose to take (chose to defect) in the original social dilemma by those who chose not to take (cooperate) in it – when retaliation is possible as this situation most closely resembles the cooperation problems faced in the real world. We are not the first to study social dilemmas with punishment and counter-punishment; Nikiforakis (2008) found that the possibility of counter-punishment reduced the willingness of cooperators to punish free riders in a public good game. In spite of this finding, we predict that Southerners in this situation will have a higher punishment rate than Non-Southerners. We believe that a similar proportion of both Southerners and Non-Southerners will punish for prosocial reasons out of a general desire to promote cooperative outcomes in future cooperation problems. However, we predict that Southerners who belong to a culture of honor have an additional motivation to punish: the restoration of their honor. Therefore, we predict Southerners will have a higher rate of punishment than Non-Southerners.

4.3 Study 1

4.3.1 Method

4.3.1.1 Subjects

We conducted the experiment using LIONESS (Aréchar, Molleman, & Gächter, subm.) where we recruited 974 subjects from the MTurk platform who were located in the US. These subjects formed 487 dyads over 19 sessions from August to October 2014. Subjects on average took 7 minutes to complete the experiment and the average payment was $1.72 (implying an hourly rate of $14.74).

The focus of this study is on the difference in frequency between Southerner and Non-Southerner first mover punishment. As is common in the literature (Cohen et al., 1996), we defined Southerners as white subjects who reported growing up in the South.[23] Nisbett & Cohen (1996) note a unique relationship between region and violence for Anglo Americans but no such relationship for other racial or ethnic categories. Therefore, we define all others as Non-Southerners. Only the subject's own geographic origin is relevant to the analysis as they played the game anonymously and as such did not know the geographic origin of their coplayer. More details of the subjects are given in the Supplementary Materials.

4.3.1.2 Procedure

Subjects reviewed experiment procedures and gave informed consent by accepting the MTurk HIT (Human Intelligence Task) before entering into LIONESS platform for the experiment. Only those who had not previously participated in a session of this study or the pilot sessions of this study[24] were allowed to proceed. Subjects then read instructions describing all stages of the experiment including the calculation of payoffs and answered control questions to ensure that they understood the rules of the game. After correctly answering

---

[23] We defined Southern states as Divisions 5, 6, and 7 of Region 3 of the US Census plus Missouri and excluding Washington, DC. More specifically, subjects were classified as Southerners if they were white and reported growing up in one of the following states: Alabama, Arkansas, Delaware, Florida, Georgia, Kentucky, Louisiana, Maryland, Mississippi, Missouri, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Virginia, or West Virginia. According to the 2010 US Census, 37.1% of the United States population lived in the South (Mackyn, Wilson, Fischetti, & Goworowska, 2011).

[24] We conducted 4 pilot sessions of Study 1 involving 97 participants. We used these pilot sessions to calibrate the payoffs in the social dilemma to ensure enough variation of cooperation and defection behavior such that we had meaningful observations of each of the four possible outcomes of Stage 1. This enabled us to analyze punishment behavior in Stages 2 for each of the four possible outcomes and answer our research question.

the control questions, subjects were placed in a virtual waiting room until either they were matched with a coplayer and placed into a dyad or they waited two minutes without being paired with a coplayer and had the option to either leave the experiment and immediately receive a $0.50 show-up fee or wait longer to be matched. Subjects remained in the same dyad with the same coplayer for the duration of the study and throughout the study their identities remained anonymous. Screenshots of the instructions and decision screens are available in the Supplementary Materials in Sections 4.6.1.1.1 and 4.6.1.1.2, respectively.

The game was played in terms of points, but payoffs were made in US Dollars at an exchange rate of 50 points equaling $1.00 USD. When a subject's coplayer left the study before it was complete, the remaining subject received their $0.50 show-up fee and we excluded the data of both from the results. At no point in the study was deception used and this study received ethical approval from the Nottingham School of Economics Research Ethics Committee.

Simultaneous Social Dilemma with Sequential Punishment. Subjects participated in a social dilemma with sequential punishment as illustrated in game tree form in Figure 4.1. Both subjects are initially endowed with 70 points and the game progresses over three stages. The first stage is a simultaneous social dilemma where both subjects choose between two options: Option A and Option B. Subjects who select Option A choose to take (choose to defect) and as such they increase their own total by 25 points and decrease the total of their coplayer by 30 points. Those who select Option B choose not to take (choose to cooperate) and as such change neither their own nor their coplayer's totals. Once both subjects make their selections, the results of this stage are shown to both subjects before the game proceeds to the next stage.

In the second stage, one of the two subjects is randomly selected and given the choice between two options: Option X and Option Y. Subjects who select Option X reduce their own total by 10 points and decrease the total of their coplayer by 30 points. Subjects who select Option Y do not change either their own or their coplayer's totals. Once this subject makes their selection, the results of this stage are shown to both subjects.

In the third stage, the other subject is given the same choice as in the second stage between the two options: Option X and Option Y. Again, subjects who select Option X reduce their own total by 10 points and decrease the total of their coplayer by 30 points. While subjects who select Option Y do not change either their own or their coplayer's totals. Once this subject makes their selection, the results of this final stage are shown to both subjects.

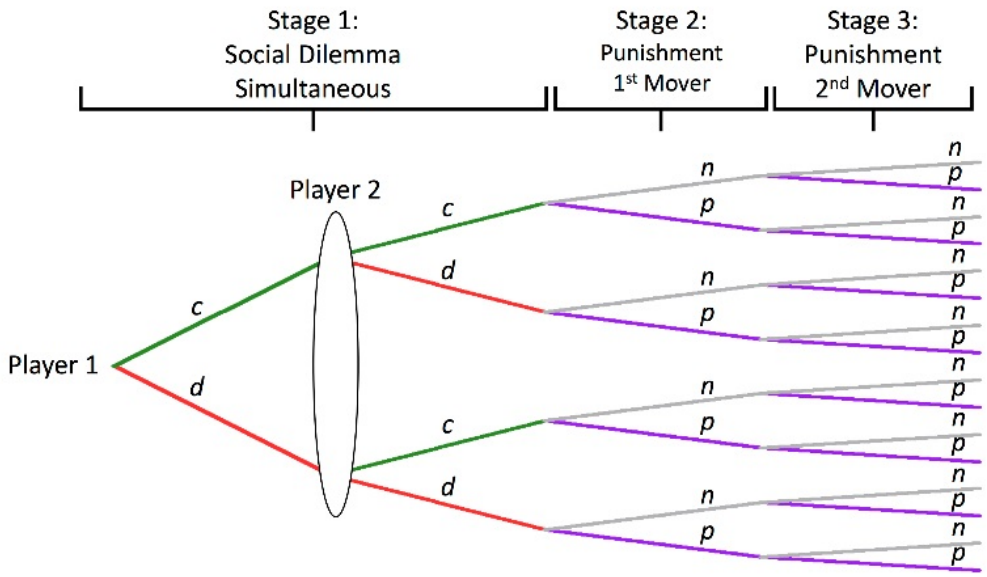


Figure 4.1. Game tree of Study 1 showing all possible outcomes. The oval which encircles the Player 2 decision nodes in Stage 1 indicates that Player 2 is unaware of Player 1's decisions before taking her own decisions. In Stage 1, the letter *c* (cooperate) refers to choosing not to take in Option B and the letter *d* (defect) refers to choosing to take in Option A. In Stages 2 and 3, the letter *n* (not punish) refers to Option Y and the letter *p* (punish) refers to Option X.

Subjects then complete a questionnaire with social and demographic questions including items about their geographic associations and ethnicity. Once they finish the questionnaire, they are given a completion code which they can enter in the MTurk website to complete the HIT and receive payment for the realized outcomes of the game.

4.3.2 Results and Discussion

4.3.2.1 Results

Subjects. After all of the 974 subjects participated in the Stage 1 simultaneous social dilemma, half were randomly selected as first mover punishers in Stage 2. Just under one third of the first mover punishers were Southerners (152 Southern and 335 Non-Southern). In Stage 3, 27% of the second mover punishers were Southerners (133 Southern and 354 Non-Southern).

Simultaneous Social Dilemma. Overall, we observe that 59.7% of subjects cooperate or chose not to take in the simultaneous social dilemma with no statistically significant difference in the rates of cooperating between Southerners (60.4%) and Non-Southerners (59.4%)($\chi^2(1) = 0.08,\ p = 0.78$). Following the simultaneous social dilemma, there were four possible outcomes for each dyad: “Self Cooperate, Other Cooperate”; “Self Cooperate, Other Defect”; “Self Defect, Other Cooperate”; and “Self Defect, Other Defect”.

Our primary focus of interest is the difference in punishment rates between Southerners and Non-Southerners who were in the “Self Cooperate, Other Defect” category. As such, we collapse the latter three categories together into one category we call “Other Outcomes.” We see that overall 26.1% of our subjects are in the “Self Cooperate, Other Defect” category with no statistically

significant difference in the rates that Southerners (27.6%) and Non-Southerners (25.4%) were in it ($\chi^2(1) = 0.40,\ p = 0.53$).

First Mover Punishment. Figure 4.2 shows the first-mover punishment rates of our subjects in both categories split by whether the first-mover punishers were Southerners or Non-Southerners.

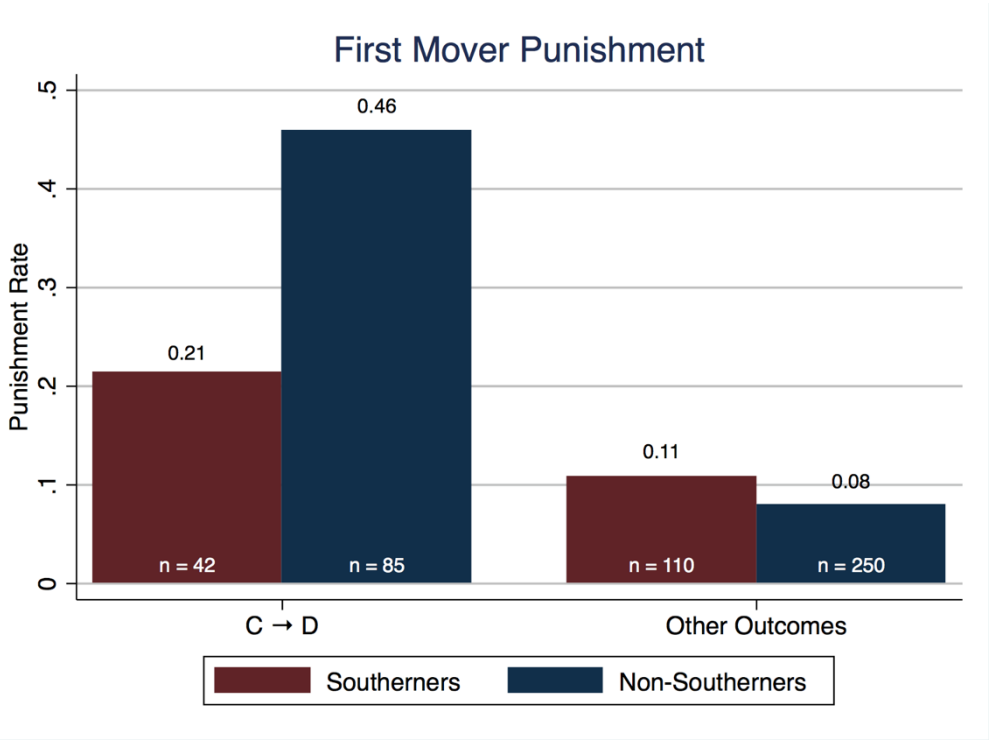


Figure 4.2. First mover punishment rates for Southerners and Non-Southerners. We show the frequencies of First-Mover Punishment in both the "Self Cooperate, Other Defect" and "Other Outcomes" categories. C → D refers to the subjects who chose to not take while their coplayers chose to take in the Stage 1 social dilemma (that is subjects in the "Self Cooperate, Other Defect" category).

The punishment rate of Southerners who chose not to take while their coplayers chose to take in the Stage 1 social dilemma (that is, Southerners in the "Self Cooperate, Other Defect" category) is 21.4% (95% confidence interval, or CI = [8.5, 34.4]). The punishment rate for Non-Southerners was 45.9% (95% CI = [35.1, 56.7]). In the "Self Cooperate, Other Defect" category, Southerners punished with statistically significant *less frequency* than

Non-Southerners ($\chi^2(1) = 7.15,\ p = 0.007$). This difference is in the opposite direction than what we predicted. In contrast, there is no statistically significant difference in the frequency of punishment behavior between Southerners and Non-Southerners in the "Other Outcomes" category ($\chi^2(1) = 0.80,\ p = 0.372$).[25] In this category Southerners punished at a rate of 10.9% ($95\%\ \text{CI} = [5.0, 16.8]$) and Non-Southerners punished at a rate of 8.0% ($95\%\ \text{CI} = [4.6, 11.4]$).

Second Mover Punishment. It is difficult to make predictions regarding second mover punishment behavior taking into account first mover punishment as it is not obvious what impact the culture of honor hypothesis has on these more extensive gameplay histories. Nevertheless, we analyze this second mover punishment behavior using a linear probability model in Table 4.1. The omitted category in this model is the "Self Cooperate, Other Defect" category of the Stage 1 social dilemma.

Not surprisingly, we observe that subjects who were punished by their coplayers in Stage 2 were statistically significantly more likely to exercise second mover punishment in Stage 3 irrespective of whether they were Southerners or Non-Southerners. Further, we see that both Southerners and non-Southerners who were in either the "Self Defect, Other Cooperate" or the "Self Cooperate, Other Cooperate" categories in the Stage 1 social dilemma were less

[25] There was no statistically significant difference in the frequency of punishment behaviour between Southerners and Non-Southerners in each of the three categories which were collapsed to form the Other Outcomes category as well. Further details are given in the Supplementary Materials, Section 4.6.1.2.2.

likely to exercise second mover punishment in Stage 3 compared to subjects in the "Self Cooperate, Other Defect" category in the Stage 1 social dilemma.

	Stage 3: Second Mover Punishment
Stage 1: Social Dilemma Game Play	
"Self Defect, Other Defect"	-0.081
	(0.078)
"Self Defect, Other Cooperate"	-0.215***
	(0.066)
"Self Cooperate, Other Cooperate"	-0.272***
	(0.057)
Culture of Honor and Interaction Terms	
Southerner	0.143
	(0.108)
Southerner X "Self Defect, Other Defect"	-0.293**
	(0.134)
Southerner X "Self Defect, Other Cooperate"	-0.148
	(0.131)
Southerner X "Self Cooperate, Other Cooperate"	-0.112
	(0.115)
Stage 2: First Mover Punishment	
Punished by Coplayer in Stage 2	0.233***
	(0.073)
Southerner X Punished by Coplayer in Stage 2	-0.068
	(0.125)
Constant	0.301***
	(0.055)
N	487
$F(9, 477)$	8.12***
R^2	0.153

Table 4.1. Second Mover Punishment Behavior in Study 1. Linear probability model where the dependent variable is a dummy variable for Second Mover Punishment. All explanatory variables are dummy variables coded 1 if true and 0 if false. The base for the Stage 1 Social Dilemma Game Play variables is the "Self Cooperate, Other Defect" category and the other categories should be interpreted as the effect in comparison to this base. Robust standard errors are given in parentheses. * p<0.10, ** p<0.05, *** p<0.01.

In general, we find that being a Southerner did not affect second mover punishment in Stage 3 ($F(5, 477) = 1.27, p = 0.274$). The one exception is that Southerners who, in Stage 1, had defected when their coplayers also

defected were statistically significantly less likely to exercise second mover punishment in Stage 3 than Non-Southerners with the same history of gameplay in the social dilemma ($F(1, 477) = 3.37, p = 0.067$).

4.3.2.2 Discussion

We found one strong difference between Southerners and Non-Southerners in Study 1: the frequency by which first mover Southerners in the "Self Cooperate, Other Defect" category punished their coplayers in Stage 2 was statistically significantly *lower* than that of Non-Southerners. This finding was exactly opposite our prediction. Beyond this unexpected finding, we observe few differences between the Southerners and Non-Southerners in Study 1. Their rates of cooperation in the social dilemma are similar. Their frequency of punishment in many cases does not vary regardless of whether they are first and second mover punishers.

Recall that punishment by a subject who chose not to take (that is, a cooperator) of a subject who chose to take (that is, a defector) in the social dilemma both reduces the inequality between the two subjects and is also the cooperator's only means of expressing anything to her coplayer. However, it is risky to punish as a first mover because the defector has the opportunity to retaliate immediately thereafter in the following stage. In spite of the possibility of retaliation, we expect a large proportion of cooperators would be motivated to exercise prosocial punishment against their defecting coplayers in this stage.

However, we further predicted that Southerners would have an additional motivation to punish. The frame of the social dilemma as a stealing game was designed to reflect the socioecological environment which gives rise to the

culture of honor and the taking action of defection was designed to challenge the honor of players from cultures of honor in addition to being a real material loss. So in addition to the standard prosocial reasons to punish, we predicted Southerners would also punish to show their willingness to defend themselves shaking off any suggestion that they could easily be taken advantage of and reestablishing their precedence in the social standing. As such, we predicted that Southerners would punish with greater frequency than Non-Southerners.

What can explain the surprising result that as first mover punishers Southern cooperators punished defectors less frequently than their Non-Southern counterparts? One likely explanation is that Southerners in this study are avoiding escalating the conflict and further exacerbating their losses because of the potential of retaliation by their coplayers in the next stage. It is often assumed that the equilibrium that exists in a culture of honor society is one in which challenges to one's honor requires a swift and reciprocal responses and in fact this is the equilibrium we describe above. However, this equilibrium is a dystopian one more associated with the seemingly endless cycles of retributive violence racking inner cities or motorcycle gangs. It stands in stark contrast to the commonly accepted view of a genteel and refined US South.

Recognizing this Cohen, Vandello, Puente, & Rantilla (1999) and Cohen & Vandello (2004) describe how there might be multiple equilibriums that result from a culture of honor and describe an alternative equilibrium where culture is predominantly characterized by politeness and cooperation. In this equilibrium, Reed writes that Southerners still hold "…that there *are* some things worth fighting for" (1982) and it is only these significant provocations that pass the threshold for honor restoring aggression. Minor provocations might politely be

ignored because Southerners are familiar with the destructive excesses of revenge and vendetta. In many ways, this equilibrium describes better the US South than the dystopian alternative described above.

Returning to the social dilemma in this study, it is possible that Southern first mover punishers who chose not to take in the social dilemma (cooperators) viewed the taking choice of their coplayers to take as a fairly minor provocation from an unknown yet seemingly belligerent individual not particularly worth fighting for in light of the immediate opportunity for retaliation. Taking in the social dilemma, while certainly an annoyance, did not pass the threshold necessitating action in light of the possibility of retaliation. Choosing not to punish these unknown yet seemingly belligerent defector coplayers, our polite Southern first movers exercised a strategic restraint compared to their Non-Southerner counterparts. Removing the possibility of retaliation might lower Southerner's threshold of response to minor offenses like defection.

In light of this unexpected finding, which is the opposite to our initial prediction, and this alternative theory, we consider the possibility that Southerners reacted differently than Non-Southerners due to the possibility of retaliation. We report below a follow-up study to test our new predictions that it is the possibility of retaliation which lowers the punishment rate of defectors by Southerners compared to Non-Southerners. We predict that, freed from the possibility of retaliation, Southern cooperators will be more likely to punish defectors.

4.4 Study 2

In Study 2, we test directly the explanation that the possibility of retaliation caused Southerners to punish defectors at a lower rate than Non-Southerners by eliminating the second mover punishment stage. Without a second mover punishment stage, retaliation against first mover punishers becomes impossible. We predict that if the possibility of retaliation is what lowered the rate at which Southern cooperators punished in Study 1, then there should be no difference in the punishment rates of defectors by Southern and Non-Southern cooperators when retaliation is impossible.

4.4.1 Methods

4.4.1.1 Subjects

We again conducted the experiment using LIONESS where we recruited 1000 subjects from MTurk who were located in the US excluding those who had participated in piloting, earlier sessions of this study, or Study 1 sessions. These subjects formed 500 dyads over 6 sessions in August 2015. The average length of session in the study was approximately 15 minutes and the average payment was $2.21, implying an hourly wage of $8.84.[26]

We define Southerners in the same way as in Study 1: white subjects who reported growing up in the South. All others were defined as Non-Southerners. Again subjects played the game anonymously so only the player's own

[26] The length of Study 2 was longer than Study 1 in spite of removing one stage because we added additional questionnaires at the end of the study.

geographic origin is relevant. More details of the subjects are given in the Supplementary Materials.

4.4.1.2 Procedure

The procedures for Study 2 follow those of Study 1 explained above in Section 4.3.1.2 with one key exception: the elimination of Stage 3, Second Mover Punishment. All other changes are detailed below.

Simultaneous Social Dilemma with Punishment by One Randomly Selected Punisher. As in Study 1, subjects participated in a social dilemma. However, in this study only one of the two subjects in the dyad was able to punish. Study 2 is illustrated in game tree form in Figure 4.3. Stage 1 in Study 2 was similar to Stage 1 in Study 1. Both subjects were endowed with 70 points and had the choice between two options in a simultaneous social dilemma: Option A and Option B. Option A meant that subjects chose to take (chose to defect) and increased their own total by 25 points and decreased their coplayers total by 35 points in Study 2.[27] Option B meant subject chose not to take (chose to cooperate) and neither their own nor their coplayer's totals were changed. After both subjects made their choices, the results of this stage were shown to both members of the dyad before proceeding to the next stage.

[27] The Option A decrease in their coplayers total changed from 30 points in Study 1 to 35 points in Study 2. We found it necessary during the piloting of Study 2 to make this change in order to achieve comparable cooperation rates in the Stage 1 social dilemma to those in Study 1.

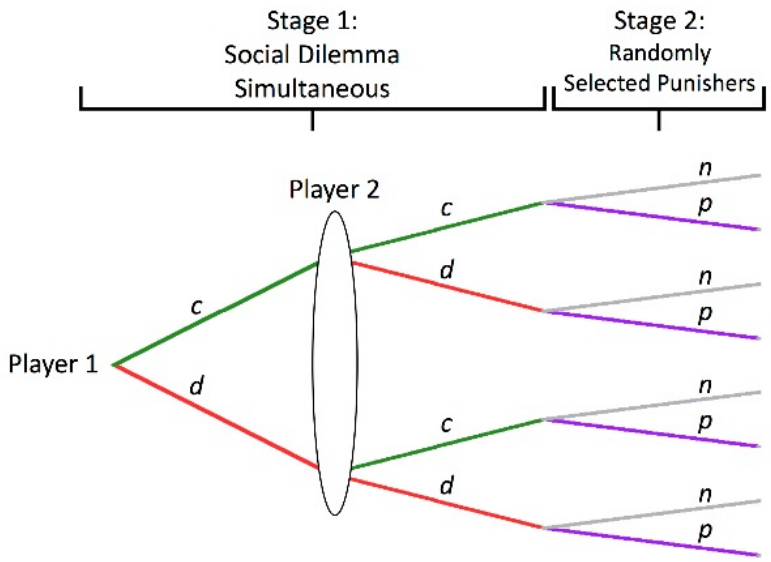


Figure 4.3. Game tree of Study 2 showing all possible outcomes. The oval which encircles the Player 2 decision nodes in Stage 1 indicates that Player 2 is unaware of Player 1's decisions before taking her own decisions. In Stage 1, the letter *c* (cooperate) refers to choosing not to take in Option B and the letter *d* (defect) refers to choosing to take in Option A. In Stage 2, the letter *n* (not punish) refers to Option Y and the letter *p* (punish) refers to Option X.

In the second stage, one of the two subjects was randomly selected and given the choice between two options: Option X and Option Y. Selecting Option X caused a reduction of that subject's own total by 10 points and of their coplayer's total by 30 points. Subjects who selected Option Y made no change to either their own or their coplayer's totals. Once this choice had been made, the results of this stage were shown to both subjects.

Subjects then completed a variety of questionnaires including social demographic questions (including information about their ethnicity and geographic associations). Once they finished the questionnaires, they were given a completion code which they could enter into the MTurk to complete the HIT and receive payment. Screenshots of the instructions and decision screens are available in the Supplementary Materials in Sections 4.6.1.1.3 and 4.6.1.1.4.

4.4.2 Results and Discussion

4.4.2.1 Results

Subjects. In Study 2, there were 1,000 subjects overall composed of 245 Southerners and 755 Non-Southerners. After all the subjects participated in the Stage 1 simultaneous social dilemma, half were randomly selected as punishers in Stage 2. Of these randomly selected first mover punishers, just under a quarter were Southerners (115 Southern and 385 Non-Southern).

Simultaneous Social Dilemma. In Study 2, we observe cooperation rates of 55.1% in the simultaneous social dilemma with no statistically significant difference in the rates at which Southerners (53.5%) and Non-Southerners (55.6%) Cooperate $(\chi^2(1) = 0.35,\ p = 0.56)$. There is again sufficient variation in not taking and taking behavior observed among our subjects to answer our research question.

We find 24.3% of subjects are in the pivotal "Self Cooperate, Other Defect" category with no statistically significant difference in the rate at which Southerners (24.5%) or Non-Southerners (24.2%) are present $(\chi^2(1) = 0.01,\ p = 0.94)$. There is no significant difference in the rates at which Southerners and Non-Southerners are present in the other three categories, but given that we have no specific predictions regarding the punishment behavior of randomly selected punishers for any of these categories we collapse them into one "Other Outcomes" category like in Study 1.

Randomly Selected Punishment. We see the punishment behavior of the randomly selected punishers in Figure 4.4. We observe no statistically significant difference in punishment rate between Southerners and Non-

Southerners in either the "Self Cooperate, Other Defect" or the Other Outcomes categories.

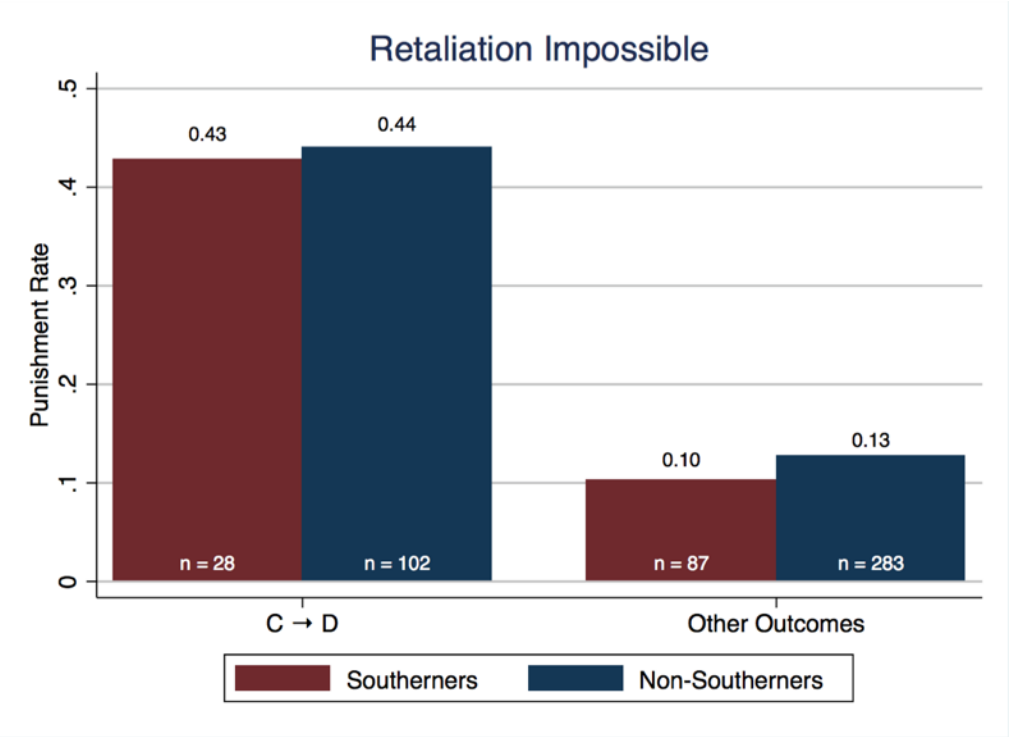


Figure 4.4. Punishment rates when retaliation is impossible for Southerners and Non-Southerners. We show the frequencies of punishment exercised by Randomly Selected Punishers in both the "Self Cooperate, Other Defect" and "Other Outcomes" categories. C → D refers to the subjects who chose to not take while their coplayers chose to take in the Stage 1 social dilemma (that is subjects in the "Self Cooperate, Other Defect" category).

In this study, the Stage 2 punishment rate of Southerners who in the Stage 1 social dilemma chose not to take punished their coplayers who chose to take (that is, Southerners in the "Self Cooperate, Other Defect" category) is 42.9% ($95\%\ \text{CI} = [23.3, 62.4]$) while the punishment rate of Non-Southerners in this same category is 44.1% ($95\%\ \text{CI} = [34.3, 53.9]$). There was no significant difference between these two punishment rates ($\chi^2(1) = 0.01,\ p = 0.905$); that is, Southerners and Non-Southerners who chose not to take in the Stage 1 social dilemma punished their coplayers who chose to take at the same rates. Likewise, in the "Other Outcomes" category, there was no statistically

significant difference in the rate at which Southerners and Non-Southerners punished ($\chi^2(1) = 0.35,\ p = 0\ .553$).[28] Southerners punished at a rate of 10.3% (95% CI $= [3.8, 16.9]$) and Non-Southerners punished at a rate of 12.7% (95% CI $= [8.8, 16.6]$) in this category.

4.4.2.2 Discussion

In Study 2, we do not observe any statistically significant differences between Southerners and Non-Southerners in the rates at which they chose not to take or chose to take in the Stage 1 simultaneous social dilemma. Further, we do not observe any statistically significant differences in the rate at which Southerners or Non-Southerners who were randomly selected choose to exercise punishment in Stage 2.

Removing the third stage – the second mover punishment stage – had a dramatic effect on the punishment rates of Southerners and decoupled the rates of punishment from geography. In Study 2, we find that Southerners do not punish at rates statistically significantly different from Non-Southerners in this simultaneous social dilemma with punishment by one randomly selected punisher.

[28] As in Study 1, there are no statistically significant differences in the rates of punishment between Southerners and Non-Southerners in the three categories collapsed to form the "Other Outcomes" category. Further details are given in Supplemental Materials, Section 4.6.1.2.5.

4.5 General Discussion

The aim of this paper is to reexamine the impact of the culture of honor in the US and in particular to see how it impacts the punishment behavior of US Southerners. The most striking finding of Study 1 is that contrary to our predictions US Southerners who chose not to take in the Stage 1 social dilemma (that is, cooperators) punish coplayers who choose to take (that is, defectors) less frequently than their Non-Southern counterparts. In light of this unexpected result, we conjectured that it was the possibility of retaliation that lowered the rate at which US Southerners were punishing. Removing the possibility of retaliation in Study 2 removed the difference in punishment rates between Southerners and Non-Southerners.

While removing the possibility of retaliation closed the gap between the punishment rates, our original prediction was that Southerners would have an additional motivation beyond Non-Southerners to punish. We predicted that Southerners would punish to defend their honor and show that they were not the sort of people who could easily be taken advantage of. We thought Southerners would exercise punishment in spite of the possibility of retaliation in order to reestablish their precedence in social standing. This is evidently not the case in the social dilemma studied in this paper.

Our findings seemingly stand in contrast to the findings of Cohen et al. in their social psychology experiments (1996). In their experiments of male university students, the responses of Southerners who were bumped into and insulted publicly by confederates of the experimenters strongly varied from those of Non-Southerners. The Southerners were more likely to perceive insult

as a threat to their masculinity, respond physiologically with increased cortisol and testosterone levels indicating greater anger and aggressiveness respectively, and entertain aggression in thought and deed in response to honor related stimuli.

However, there were some caveats to these findings. First, only publicly insulted Southerners were more likely to believe that others looked down on them. Southerners who were insulted privately were unlikely to believe others looked down on them and their beliefs in this regard were no different from the control groups who were not insulted. Second, Cohen et al. observed that Southerners who were insulted were no more likely than Non-Southerners to project aggression onto neutral stimuli unrelated to honor. It seems that challenges to Southerners matter only to the extent that they are public and their responses to those challenges were relevant only to domains relating to honor. Southerners in our study might not have responded with honor-restoring punishment against coplayers who chose to take (that is, defectors) in the Stage 1 social dilemma because the interaction was anonymous and private and therefore not regarded as a threat to the Southerners' honor. Further, they may have viewed the Stage 2 punishment stage as a neutral – as opposed to an honor related – domain.

In another study, Leung & Cohen (2011) employed a within- and between-subject design and they observe variation in the extent to which individuals from a culture of honor exhibit positive reciprocity, negative reciprocity, and virtue. They find that Southerners from a culture of honor who most endorse the values associated with such a culture – in this instance, negative reciprocity in response to insult – were most likely to exhibit positive reciprocity in an unrelated

situation. The effect was opposite for Non-Southerners. In a second study, they found that Southerners who most endorse culture of honor values exhibited greater honesty in an unrelated task whereas again the effect was opposite for Non-Southerners. However, Southerners exhibited lower levels of honesty in this second study when the order of tasks was reversed and the task measuring honesty came before the Southerners evaluated culture of honor values. It appears that evaluating values associated with a culture of honor activates other behavior Southerners also associated with honor – in this case, honesty – and without that prime the parallel associations were less prevalent. In our study, it is possible that the culture of honor values of the Southerners were not activated in the Stage 1 social dilemma and this is why Southerners in our study do not respond to their coplayer's taking behavior with greater levels of honor restoring punishment in Stage 2.

What might explain the lower rate of Stage 2 punishment that Southerners exercised compared to Non-Southerners when retaliation was possible? These results are related to the findings of Cohen et al. (1999) in another social psychology study. In their study, a fellow participant in the study (actually a confederate of the researchers) carried out a series of escalating annoyances against Southerners and Non-Southerners who were participating in an unrelated task. Exposed to these repeated annoyances the anger of Non-Southerners steadily increased while their amusement steadily decreased over the first few annoyances until it plateaued. They contend that Non-Southerners were more skilled at communicating their displeasure through these small displays of hostility which eventually diffused their anger. In contrast, the emotional responses of Southerners to the first few annoyances were relatively

flat until they exploded with dramatic displays of hostility. Attributing this pattern of behavior to the Southern norm of politeness, Cohen et al. contend Southerners were more willing to bottle up their displeasure in the short term leading to eventual explosions of anger. This pattern might explain why Southerners punished coplayers who defect at a lower rate than Non-Southerners following the social dilemma in our study: they were exercising restraint through the norm Southern politeness.

While not present in the way initially predicted, this study has demonstrated that some variant of the culture of honor is still shaping the behavior of US Southerners. Further, this study raises many areas of possible future research. It would be interesting to manipulate the social nature of interactions in a social dilemma. Perhaps Southerners whose coplayers publicly take from them would exercise punishment at greater rates than Non-Southerners. Another interesting manipulation would be to try to prime culture of honor responses by asking subjects to evaluate values associated with a culture of honor prior to participating in the social dilemma. Perhaps Southerners who most endorse culture of honor values would be primed to exercise punishment at greater rates in a subsequent social dilemma. Finally, repeated interactions would enable us to test the idea that the norm of Southern politeness is what is restraining expressions of discontent. Perhaps the polite disregard of Southerners who are repeatedly defected against would eventually lead to an explosion of stronger punishment behavior.

In conclusion we find that prosocial punishment has different roles to play in the South and Non-South. We have demonstrated that there is variation of punishment behavior within the US and that the culture of honor still shapes the

behavior of US Southerners but in ways less straightforward than we initially predicted. We have found that the punishment rate of players who take (that is, defectors) in a social dilemma by Southerners is lower than that of Non-Southerners but only in instances when retaliation against the punisher is possible. Removing any possibility of retaliation leads to no differences in the rates of punishment between Southerners and Non-Southerners. We explain these differences as the result of a polite, yet strategic indifference to their coplayer's taking behavior on the part of Southerners in light of the possibility of retaliation. These new insights have useful applications in understanding how best to approach social problems in both the United States, but also in other societies where the culture of honor is present.

4.6 Supplementary Materials

4.6.1.1 Instructions and Decision Screens

We include copies of the instructions and screen shots of the decision screens the subjects saw during this study. Full copies of the software we used are available at request from the authors.

4.6.1.1.1 Study 1 Instructions

Instructions

In this experiment you will be playing with another MTurker who also accepted this HIT, and who is completing it **at the same time**. It is therefore important that you complete this HIT **without interruptions**. Including the time for reading these instructions, the experiment will take about 8 minutes to complete.

During this experiment, you will play a game in which you can earn Points. The number of Points you earn depends on your decisions and the decisions of the other player. At the end of the experiment your Points will be converted into real money (**50 Points = $1.00**). You will receive a code to collect your payment via MTurk upon completion.

Your task

At the beginning of the experiment, you and the other player both receive 70 Points.
You will play a game that consists of **3 Stages**.

Stage I

In Stage I, you will decide between two options:

Option A	or	**Option B**
Increase your total by **25 Points**. **Reduce** the total of the other by **30 Points**.		Do **not change** your total. Do **not change** the total of the other.

The other player makes a choice between Option A and Option B **at the same time**. When both you and the other have made your decisions, the outcome of this Stage will be shown to both players. You will then proceed to Stage II.

Stage II

In Stage II, one player (you or the other) is randomly chosen to make a decision. If you are chosen, you will decide between two options:

Option X	or	**Option Y**
Reduce your total by **10 Points**. **Reduce** the total of the other by **30 Points**.		Do **not change** your total. Do **not change** the total of the other.

Once this decision has been made, the outcome of this Stage will be shown to both players. You will then proceed to Stage III.

Stage III

In Stage III, the roles are reversed. If you did not make decision in Stage II, you will decide between Options X and Option Y.

Once this decision has been made, the outcome of this Stage will be shown to both players. This ends the game. A brief questionnaire concludes this HIT.

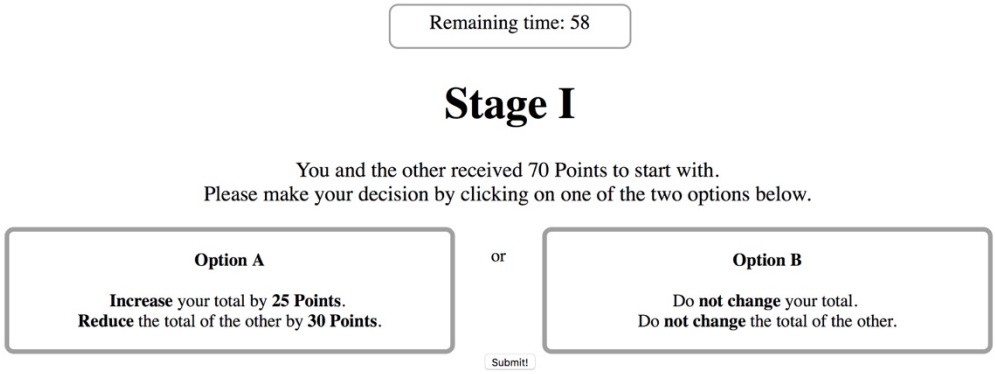
Note: Please make your decisions within the **time limit** shown on your screen. If you fail to do so, you will be **removed from the experiment** and we will **not** be able to pay you.

Please click the link below if you understood the instructions. A brief check will follow to test your understanding of the experiment.

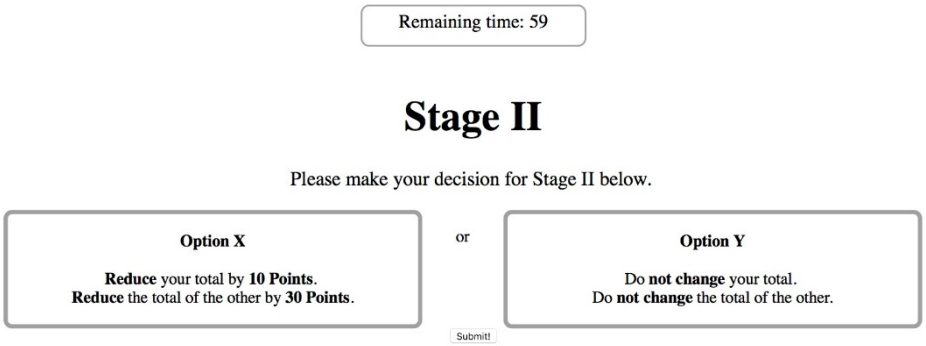
I have read and understood the instructions. Continue!

4.6.1.1.2 Study 1 Decision Screens

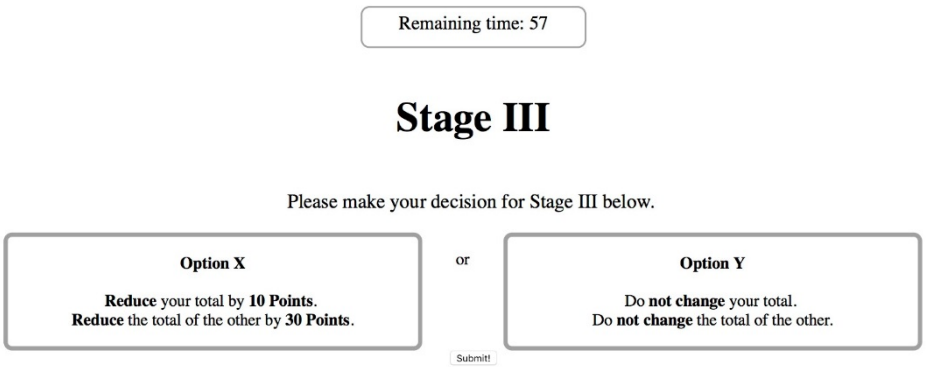
Stage 1 Decision Screen



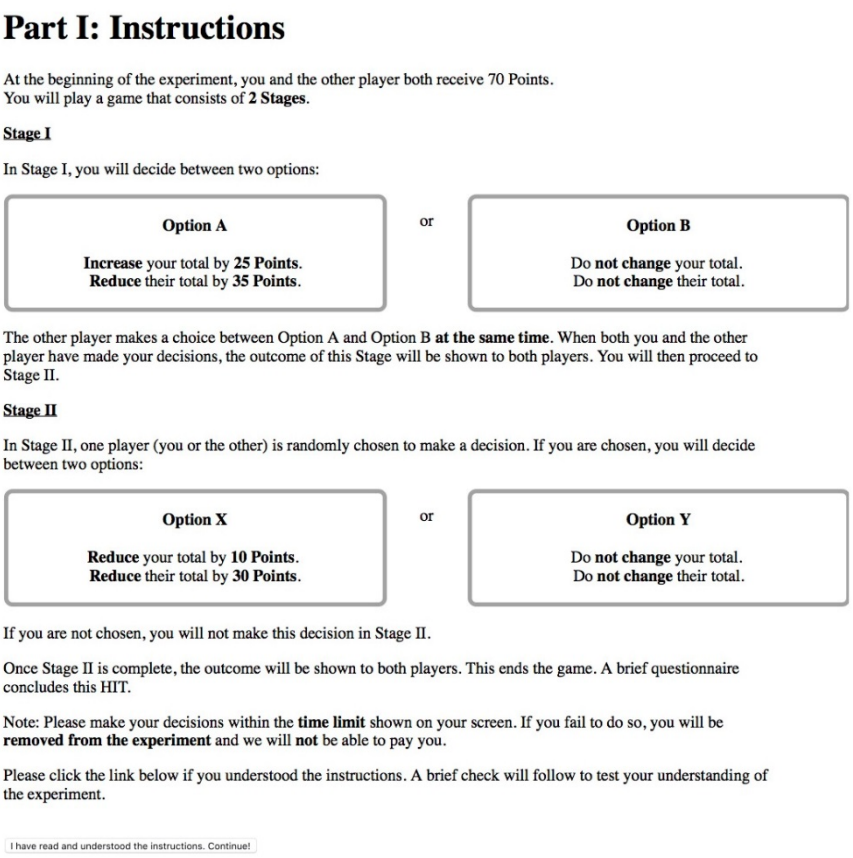
Stage 2 Decision Screen



Stage 3 Decision Screen



4.6.1.1.3 Study 2 Instructions



Part I: Instructions

At the beginning of the experiment, you and the other player both receive 70 Points.
You will play a game that consists of **2 Stages**.

Stage I

In Stage I, you will decide between two options:

Option A	or	**Option B**
Increase your total by **25 Points**. **Reduce** their total by **35 Points**.		Do **not change** your total. Do **not change** their total.

The other player makes a choice between Option A and Option B **at the same time**. When both you and the other player have made your decisions, the outcome of this Stage will be shown to both players. You will then proceed to Stage II.

Stage II

In Stage II, one player (you or the other) is randomly chosen to make a decision. If you are chosen, you will decide between two options:

Option X	or	**Option Y**
Reduce your total by **10 Points**. **Reduce** their total by **30 Points**.		Do **not change** your total. Do **not change** their total.

If you are not chosen, you will not make this decision in Stage II.

Once Stage II is complete, the outcome will be shown to both players. This ends the game. A brief questionnaire concludes this HIT.

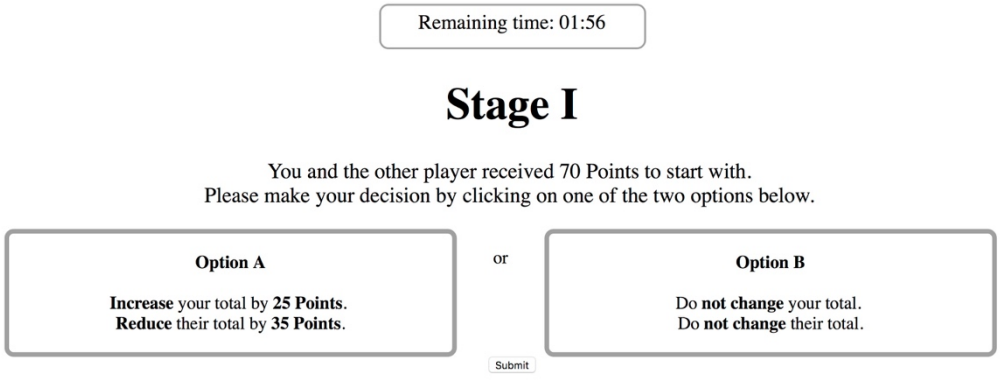
Note: Please make your decisions within the **time limit** shown on your screen. If you fail to do so, you will be **removed from the experiment** and we will **not** be able to pay you.

Please click the link below if you understood the instructions. A brief check will follow to test your understanding of the experiment.

I have read and understood the instructions. Continue!

4.6.1.1.4 Study 2 Decision Screens

Stage 1 Decision Screen



Remaining time: 01:56

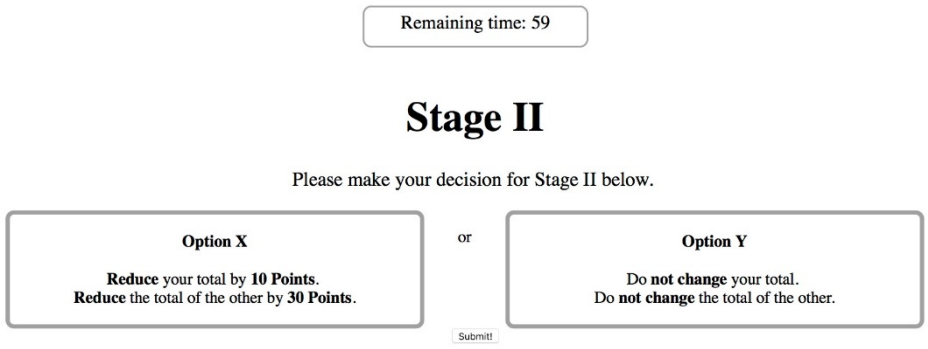
Stage I

You and the other player received 70 Points to start with.
Please make your decision by clicking on one of the two options below.

Option A	or	**Option B**
Increase your total by **25 Points**. **Reduce** their total by **35 Points**.		Do **not change** your total. Do **not change** their total.

Submit

Stage 2 Decision Screen



4.6.1.2 Supporting Analysis

4.6.1.2.1 Study 1 Subject Pool Statistics

In Table 4.2, we list social demographic characteristics as reported in Study 1 by subjects split by whether they were first and second mover punishers. We provide comparisons between our Southern and Non-Southern subjects.

			Age			
	N	Female	Years	SD	Min	Max
			First Mover Punishers			
Southerners	152	43%	32.6	10.8	19	67
Non-Southerners	335	43%	33.0	10.8	18	75
			Second Mover Punishers			
Southerners	133	49%	32.1	9.9	19	66
Non-Southerners	354	45%	32.5	10.1	19	65

Table 4.2. Social demographic characteristics as reported in Study 1 split by whether subjects were first and second mover punishers.

4.6.1.2.2 Study 1 First Mover Punishment Rates for All Categories

As reported in Section 4.3.2.1, there was only one category with statistically significant difference in the frequency of punishment between Southerners and Non-Southerners: the "Self Cooperate, Other Defect" category. Therefore, we collapsed the remaining three categories together in Figure 4.2. Here in Figure

4.5 we separate out each of these categories: “Self Defect, Other Defect”; “Self Defect, Other Cooperate”; and “Self Cooperate, Other Cooperate”. We further report the 95% confidence intervals for each category.

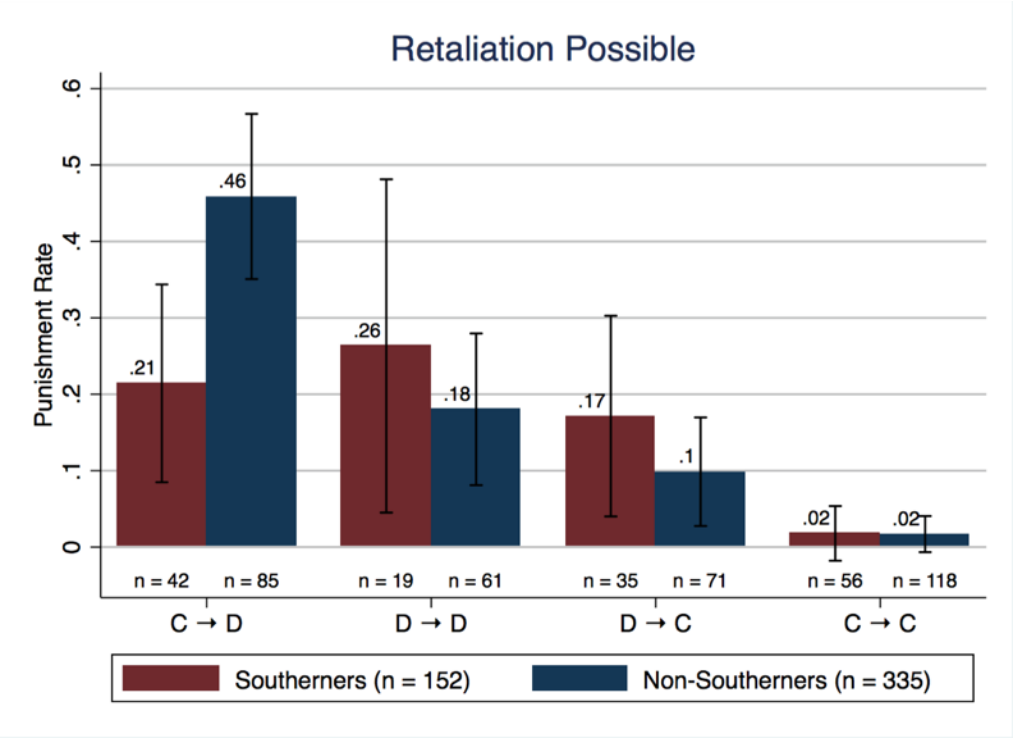


Figure 4.5. First Mover punishment rates when retaliation is possible for Southerners and Non-Southerners. C → D refers to the “Self Cooperate, Other Defect” category, D → D refers to the “Self Defect, Other Defect” category, D → C refers to the “Self Defect, Other Cooperate” category, and C → C refers to the “Self Cooperate, Other Cooperate” category. The error bars show the 95% confidence intervals.

4.6.1.2.3 Tukey-Kramer Pairwise Comparisons of Punishment Outcomes

In addition to our probit regression analysis of second mover punishment we also conduct Tukey-Kramer Pairwise Comparisons of means. This multiple comparison procedure compares the means of different categories accounting for multiple comparisons and the null hypothesis are that all the means being compared are from the same population. We separate this analysis by whether our subjects are Southerners in Table 4.3 or Non-Southerners in Table 4.4.

There is a statistically significant difference in frequency of second mover punishment of the "Self Cooperate, Other Defect" category compared to the other categories for our Southern subjects.

	C → D (0.467)	D → C (0.152)	C → C (0.068)
D → D (0.115)	0.3513***	0.0361	0.0472
C → D (0.467)	-	0.3152***	0.3985***
D → C (0.152)	-	-	0.0833

Table 4.3. Tukey-Kramer Pairwise Pairwise Comparisons of the Different Categories of Second Mover Punishment for Southerners. Means by category are listed in parenthesis. Difference of means are given. D → D refers to the "Self Defect, Other Defect" category, C → D refers to the "Self Cooperate, Other Defect" category, D → C refers to the "Self Defect, Other Cooperate" category, and C → C refers to the "Self Cooperate, Other Cooperate" category. Significance of the Tukey-Kramer test are: * $p<0.10$, ** $p<0.05$, *** $p<0.01$.

The statistically significant difference in frequency of second mover punishment of the "Self Cooperate, Other Defect" category compared to the other categories is not present for our Non-Southern subjects. Rather for our Non-Southern subjects, those in the "Self Cooperate, Other Cooperate" category were much less likely to exercise second mover punishment than those in other categories.

	C → D (0.329)	D → C (0.170)	C → C (0.031)
D → D (0.259)	0.0697	0.0890	0.2285***
C → D (0.329)	-	0.1587**	0.2982***
D → C (0.170)	-	-	0.1394**

Table 4.4. Tukey-Kramer Pairwise Pairwise Comparisons of the Different Categories of Second Mover Punishment for Non-Southerners. Means by category are listed in parenthesis. Difference of means are given. D → D refers to the "Self Defect, Other Defect" category, C → D refers to the "Self Cooperate, Other Defect" category, D → C refers to the "Self Defect, Other Cooperate" category, and C → C refers to the "Self Cooperate, Other Cooperate" category. Significance of the Tukey-Kramer test are: * p<0.10, ** p<0.05, *** p<0.01.

4.6.1.2.4 Study 2 Subject Pool Statistics

In Table 4.5, we list social demographic characteristics as reported in Study 2 by subjects split by whether they were first and second mover punishers. We provide comparisons between our Southern and Non-Southern subjects.

			Age			
	N	Female	Years	SD	Min	Max
			Randomly Selected Punishers			
Southerners	115	49%	35.9	11.8	20	77
Non-Southerners	385	48%	34.0	11.4	18	69
			Randomly Selected Non-Punishers			
Southerners	130	58%	34.4	11.0	18	68
Non-Southerners	370	49%	34.9	11.3	19	72

Table 4.5. Social demographic characteristics as reported in Study 2 split by whether subjects were first and second mover punishers.

4.6.1.2.5 Study 2 First Mover Punishment Rates for All Categories

As reported in Section 4.4.2.1, we did not observe any statistically significant difference in the frequency of punishment between Southerners and Non-Southerners in the "Self Cooperate, Other Defect" category when retaliation is

impossible. Following the convention of Figure 4.2, we nonetheless collapsed the remaining three categories together in Figure 4.4. Here in Figure 4.6 we separate out each of these categories – "Self Defect, Other Defect"; "Self Defect, Other Cooperate"; and "Self Cooperate, Other Cooperate" – and further report the 95% confidence intervals for each category.

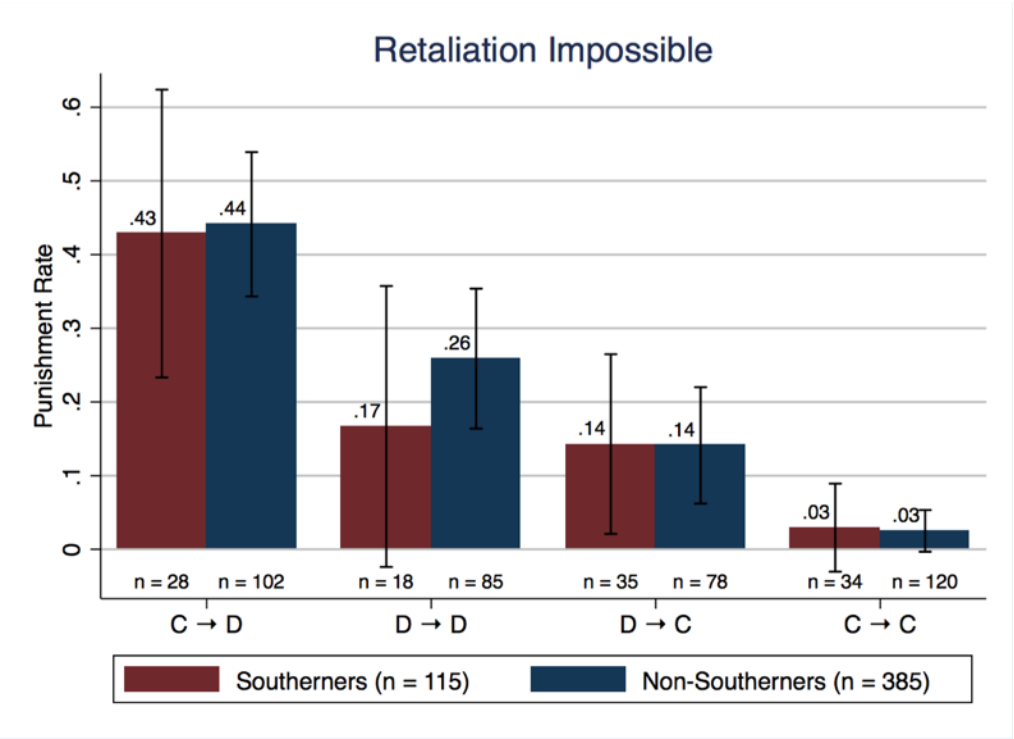


Figure 4.6. Punishment rates when retaliation is impossible for randomly selected Southern and Non-Southern punishers. C → D refers to the "Self Cooperate, Other Defect" category, D → D refers to the "Self Defect, Other Defect" category, D → C refers to the "Self Defect, Other Cooperate" category, and C → C refers to the "Self Cooperate, Other Cooperate" category. The error bars show the 95% confidence intervals.

Chapter 5

Conclusions

This thesis is an examination of social preferences using experimental economics techniques to explore specific research questions. We draw our results and conclusions from data collected using laboratory experiments in the United Kingdom and Turkey as well as data we have collected online using innovative research software and platforms. We view the diversity of subjects in our studies as a strength and as a significant (if limited) response to concerns regarding the generalizability of experimental economics research. We analyze data from incentivized tasks including modified dictator and ultimatum games, public good games, and novel social dilemmas as well as data from non-incentivized tasks including psychological measures and indicators of satisfaction in the outcomes of vignettes. We believe that data collected from a mixture of methods offers more complete answers to our research questions and greater insight to the phenomena we are investigating. The result of this research is three studies on social preferences.

In Chapter 2, we elicit stated and revealed inequality aversion in three subject pools, replicating the work of Loewenstein et al. (1989) and Blanco et al. (2011) in doing so. To elicit stated inequality aversion, we follow the design of the Loewenstein et al. (1989) tasks in which subjects read several scenarios and ranked their satisfaction with a variety of outcomes. From these rankings, we estimate individual "social utility" functions with the form of the Fehr and Schmidt (1999) model. To elicit revealed inequality aversion, we follow the design of Blanco et al. (2011) in which individual-level parameters of advantageous and disadvantageous inequality aversion are inferred from their choices in the *Modified Dictator* and *Ultimatum* games, respectively. We find that stated inequality aversion is qualitatively similar to Loewenstein et al. (1989) in each of our subject pools, but that there are some notable differences between our findings on revealed inequality aversion and those of Blanco et al. (2011). We also find revealed aversion to advantageous inequality to be stronger than revealed aversion to disadvantageous inequality for a majority of our subjects, contrary to a core assumption of the Fehr and Schmidt (1999) model of inequity aversion. Utilizing our within subject design, we find only weak correlation between the stated and revealed preferences of our subjects.

Our finding that stated and revealed preferences are only weakly correlated (and have some distinct properties) suggests that the scenario tasks and the laboratory games of Chapter 2 are measuring (at least somewhat) different things, for example attitudes in one case and willingness to act on them by making particular trade-offs in another. To the extent that this interpretation of our data is correct and generalizes, the finding would have important implications for studies that consider only revealed, or only stated, preference

techniques. Consequently, further study of the relationship between these two modes of social preference would be useful.

In Chapter 3, we consider whether the inequality aversion model of Fehr and Schmidt (1999) explains free riding and conditional cooperation. The model makes the striking prediction that a sufficiently inequality averse player will exactly match the contribution of their coplayer in a two-player linear public good game; and that any player whose inequality aversion is insufficient will free ride. We test this by eliciting advantageous inequality aversion using the Blanco et al. (2011) modified dictator game and eliciting player type using a two-player Fischbacher et al. (2001) public good game played with the strategy method. Looking at the associations between these measures, we find that the predictions of the Fehr and Schmidt (1999) model account for the free riding and conditional cooperation behavior of approximately one third of our subjects. Contrary to prediction, half our subjects who are classified as free riders were predicted by the Fehr and Schmidt (1999) model, using our elicitation of their parameter of advantageous inequality aversion, to be sufficiently inequality averse to match the contribution of their coplayers. Further, the majority of our subjects are classified as imperfect conditional cooperators or other types, neither of which possibility is predicted by the Fehr and Schmidt (1999) model.

One of the reasons the Fehr an Schmidt (1999) model is so prevalently used is because of its elegant simplicity. However, a consequence of this simplicity is the non-compromising nature of its predictions. Were we able to loosen the predicted form of cooperation to include imperfect, as well as perfect, conditional cooperation, then the overall success of the model would

significantly improve. In fact, a non-linear form of inequality aversion may allow for some (but not all) forms of imperfect conditional cooperation and further study of this is warranted. Additionally, our analysis of the psychological measures alongside inequality aversion suggests that inequality aversion is one of multiple explanations of conditional cooperation. Further study of the ways in which these multiple explanations interact would be useful.

In Chapter 4, we investigated whether variation in the directionality of punishment within a social dilemma can be explained by the *Culture of Honor* hypothesis. This hypothesis states that individuals from particular cultures must respond to a challenge to their honor with violent aggression so as not to gain a reputation as someone easily taken advantage of. Subjects from across the United States of America participated online in a novel social dilemma with sequential punishment. In contrast to our prediction, we find that Southerners (who are commonly attributed as belonging to a culture of honor) punish less frequently than Non-Southerners. In the light of this unexpected result, we conjectured that the possibility that Southerners reacted differently than Non-Southerners to the possibility of retaliation. We conducted a follow-up study where we removed the possibility of retaliation. In this study, we found no difference in the frequency of punishment between Southerners and Non-Southerners.

While we observed differences between Southerners and Non-Southerners in these studies, they opposed our initial predictions based on the culture of honor hypothesis. We conclude that the possibility of retaliation has a significant impact on the punishment behavior of Southerners; that is, on subjects from a culture of honor. Further study of how Southerners respond to

the possibility of retaliation would be useful. In particular, it would be interesting to see what impact the possibility of retaliation has on cooperative and punishment behavior in repeated interactions among those from a culture of honor. Another area ripe for further study is the role of anonymity in the punishment behavior of those from a culture of honor. Behavior attributed to a culture of honor might only manifest in situations perceived to affect honor and the removal of anonymity might provoke such a response. Manipulating the visibility of decisions in the game might affect how many Southerners exercise punishment. Again, further research would be useful.

Overall these studies explore individual social preferences using the tools of experimental economics. In particular, we have studied in some depth the concepts of inequality aversion, conditional cooperation, and punishment. While adding new insights to these areas, we find that our studies raise many more research questions. We conclude that current theories of social preferences do improve on standard economic theory, but much remains to be done to understand these phenomena. It is our hope that these studies provide new insights into these areas and more importantly provide inspiration for further research.

Bibliography

Adams, J. S. (1965). Inequity In Social Exchange. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 2, pp. 267–299). New York: Academic Press. http://doi.org/10.1016/S0065-2601(08)60108-2

Aréchar, A. A., Gächter, S., & Molleman, L. (2016). Conducting interactive experiments online. *Subm.*

Balafoutas, L., Kerschbamer, R., & Sutter, M. (2012). Distributional preferences and competitive behavior. *Journal of Economic Behavior and Organization*, *83*(1), 125–135. http://doi.org/10.1016/j.jebo.2011.06.018

Barr, A., Wallace, C., Ensminger, J., Henrich, J., Barrett, C., Bolyanatz, A., … Ziker, J. (2009). Homo Æqualis: A cross-society experimental analysis of three bargaining games. *Centre for the Study of African Economies Working Paper Series*, *WPS/2009-0*(422), 1–29. http://doi.org/Documento CEDE No. 2009-09. Available at SSRN: http://ssrn.com/abstract=1485862

Benabou, R., & Tirole, J. (2006). Incentives and Prosocial Behavior. *The American Economic Review*, *96*(5), 1652–1678.

Blanco, M., Engelmann, D., Koch, A. K., & Normann, H.-T. (2010). Belief elicitation in experiments : is there a hedging problem ? *Experimental Economics*, *13*, 412–438. http://doi.org/10.1007/s10683-010-9249-1

Blanco, M., Engelmann, D., Koch, A. K., & Normann, H. T. (2014). Preferences and beliefs in a sequential social dilemma: A within-subjects analysis. *Games and Economic Behavior*, *87*, 122–135. http://doi.org/10.1016/j.geb.2014.05.005

Blanco, M., Engelmann, D., & Normann, H. T. (2011). A within-subject analysis of other-regarding preferences. *Games and Economic Behavior*, *72*(2), 321–338. http://doi.org/10.1016/j.geb.2010.09.008

Bohnet, I., & Zeckhauser, R. (2004). Trust, risk and betrayal. *Journal of Economic Behavior and Organization*, *55*(4 SPEC.ISS.), 467–484. http://doi.org/10.1016/j.jebo.2003.11.004

Bolton, G. E. (1991). A Comparative Model of Bargaining: Theory and Evidence. *The American Economic Review*, *81*(5), 1096–1136.

Bolton, G. E., & Ockenfels, A. (2000). ERC : A Theory of Equity, Reciprocity, and Competition. *American Economic Review*, *90*(1), 166–193. Retrieved from http://www.jstor.org/stable/10.2307/117286

Brown, R. P., Osterman, L. L., & Barnes, C. D. (2009). School violence and the culture of honor. *Psychological Science*, *20*(11), 1400–1405. http://doi.org/10.1111/j.1467-9280.2009.02456.x

Camerer, C. F., & Fehr, E. (2004). Measuring Social Norms and Preferences using Experimental Games: A Guide for Social Scientists. In J. Henrich, R. Boyd, S. Bowles, C. F. Camerer, E. Fehr, & H. Gintis (Eds.), *Foundations of Human Sociality: Experimental and Ethnographic Evidence from 15 Small-Scale Societies* (pp. 55–95). New York: Oxford

University Press. http://doi.org/10.1093/0199262055.003.0003

Chandler, J., Mueller, P., & Paolacci, G. (2014). Nonnaïveté among Amazon Mechanical Turk workers: consequences and solutions for behavioral researchers. *Behavior Research Methods*, *46*(1), 112–30. http://doi.org/10.3758/s13428-013-0365-7

Charness, G., & Dufwenberg, M. (2006). Promises and Partnership. *Econometrica*, *74*(6), 1579–1601. http://doi.org/10.1111/j.1468-0262.2006.00719.x

Chaudhuri, A. (2011). Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature. *Experimental Economics*, *14*(1), 47–83. http://doi.org/10.1007/s10683-010-9257-1

Cheung, S. L. (2014). New insights into conditional cooperation and punishment from a strategy method experiment. *Experimental Economics*, *17*(1), 129–153. http://doi.org/10.1007/s10683-013-9360-1

Chu, R., Rivera, C., & Loftin, C. (2000). Herding and Homicide: An Examination of the Nisbett-Reaves Hypothesis. *Social Forces*, *78*(3), 971–987. http://doi.org/10.1093/sf/78.3.971

Cohen, D., & Nisbett, R. (1994). Self-protection and the culture of honor: Explaining southern violence. *Personality and Social Psychology ...*, *20*(5), 551–567. Retrieved from http://psp.sagepub.com/content/20/5/551.short

Cohen, D., & Nisbett, R. E. (1997). Field Experiments Examining the Culture of Honor: The Role of Institutions in Perpetuating Norms about Violence.

Personality and Social Psychology Bulletin, *23*, 1188–1199.

Cohen, D., Nisbett, R. E., Bowdle, B. F., & Schwarz, N. (1996). Insult, aggression, and the southern culture of honor: an "experimental ethnography". *Journal of Personality and Social Psychology*, *70*(5), 945–959. http://doi.org/10.1037/0022-3514.70.5.945

Cohen, D., & Vandello, J. (2004). The Paradox of Politeness. In M. Anderson (Ed.), *Cultural Shaping of Violence* (pp. 119–132). West Lafayette, Indiana: Purdue University Press.

Cohen, D., Vandello, J., Puente, S., & Rantilla, A. (1999). "When You Call Me That, Smile!" How Norms for Politeness, Interaction Styles, and Aggression Work Together in Southern Culture. *Social Psychology Quarterly*, *62*(3), 257–275.

Cohen, T. R., Wolf, S. T., Panter, A. T., & Insko, C. A. (2011). Introducing the GASP scale: a new measure of guilt and shame proneness. *Journal of Personality and Social Psychology*, *100*(5), 947–66. http://doi.org/10.1037/a0022641

Croson, R. T. A. (2007). Theories of commitment, altruism and reciprocity: Evidence from linear public goods games. *Economic Inquiry*, *45*(2), 199–216. http://doi.org/10.1111/j.1465-7295.2006.00006.x

Cubitt, R., Gächter, S., & Quercia, S. (2015). *Conditional Cooperation and Betrayal Aversion* (CeDEx Discussion Paper Series No. 2015–14). Nottingham, UK. Retrieved from http://nottingham.ac.uk/cedex/documents/papers/cedex-discussion-paper-

2015-14.pdf

Dannenberg, A., Riechmann, T., Sturm, B., & Vogt, C. (2007). *Inequity aversion and individual behavior in public good games: An experimental investigation* (ZEW Discussion Paper No. 07–034). *ZEW-Centre for European Economic Research*. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=991555

Dannenberg, A., Riechmann, T., Sturm, B., & Vogt, C. (2012). Inequality aversion and the house money effect. *Experimental Economics*, *15*(3), 460–484. http://doi.org/10.1007/s10683-011-9308-2

Dariel, A., & Nikiforakis, N. (2014). Cooperators and reciprocators: A within-subject analysis of pro-social behavior. *Economics Letters*, *122*(2), 163–166. http://doi.org/10.1016/j.econlet.2013.10.033

Dufwenberg, M., Gächter, S., & Hennig-Schmidt, H. (2011). The framing of games and the psychology of play. *Games and Economic Behavior*, *73*(2), 459–478. http://doi.org/10.1016/j.geb.2011.02.003

Eckel, C. C., & Grossman, P. J. (1998). Are Women Less Selfish Than Men?: Evidence From Dictator Experiments. *The Economic Journal*, *108*(May), 726–735. http://doi.org/10.1111/1468-0297.00311

Elster, J. (1990). Norms of Revenge. *Ethics*, *100*(4), 862–885.

Engel, C. (2011). Dictator games: A meta study. *Experimental Economics*, *14*(4), 583–610. http://doi.org/10.1007/s10683-011-9283-7

Falk, A., & Fischbacher, U. (2006). A theory of reciprocity. *Games and*

Economic Behavior, *54*(2), 293–315. http://doi.org/10.1016/j.geb.2005.03.001

Fatas, E., & Mateu, G. (2015). Antisocial punishment in two social dilemmas. *Frontiers in Behavioral Neuroscience*, *9*, 107. http://doi.org/10.3389/fnbeh.2015.00107

Fehr, E., & Fischbacher, U. (2004). Social norms and human cooperation. *Trends in Cognitive Sciences*, *8*, 185–190. http://doi.org/10.1016/j.tics.2004.02.007

Fehr, E., & Gächter, S. (2000). Cooperation and Punishment in Public Goods Experiments. *The American Economic Review*, *90*(4), 980–994. http://doi.org/10.1126/science.151.3712.867-a

Fehr, E., & Schmidt, K. M. . (1999). A Theory of Fairness, Competition, and Cooperation. *The Quarterly Journal of Economics*, *114*(3), 817–868.

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, *10*(2), 171–178. http://doi.org/10.1007/s10683-006-9159-4

Fischbacher, U., & Gächter, S. (2010). Social Preferences , Beliefs , and the Dynamics of Free Riding in Public Goods Experiments. *American Economic Review*, *100*(1), 541–556.

Fischbacher, U., Gächter, S., & Fehr, E. (2001). Are people conditionally cooperative ? Evidence from a public goods experiment. *Economics Letters*, *71*, 397–404.

Fischbacher, U., Gächter, S., & Quercia, S. (2012). The behavioral validity of the strategy method in public good experiments. *Journal of Economic Psychology*, *33*(4), 897–913. http://doi.org/10.1016/j.joep.2012.04.002

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*(4), 25–42. Retrieved from http://www.jstor.org/stable/4134953

Gächter, S., Herrmann, B., & Thöni, C. (2010). Culture and cooperation. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *365*(1553), 2651–61. http://doi.org/10.1098/rstb.2010.0135

Gächter, S., & Schulz, J. F. (2016). Intrinsic honesty and the prevalence of rule violations across societies. *Nature*, *531*(7595), 1–11. http://doi.org/10.1038/nature17160

Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality*, *37*(6), 504–528. http://doi.org/10.1016/S0092-6566(03)00046-1

Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with ORSEE. *Journal of the Economic Science Association*, *1*(1), 114–125. http://doi.org/10.1007/s40881-015-0004-4

Grosjean, P. (2014). A History of Violence: the Culture of Honor and Homicide in the US South. *Journal of the European Economic Association*, *12*(5), 1285–1316. http://doi.org/10.1111/jeea.12096

Güth, W., Schmittberger, R., & Schwarze, B. (1982). An experimental analysis

of ultimatum bargaining. *Journal of Economic Behavior & …*, *3*, 367–388. Retrieved from http://www.sciencedirect.com/science/article/pii/0167268182900117

Hartig, B., Irlenbusch, B., & Kölle, F. (2015). Conditioning on what? Heterogeneous contributions and conditional cooperation. *Journal of Behavioral and Experimental Economics* , *55*, 48–64. http://doi.org/10.1016/j.socec.2015.01.001

Henrich, J., Boyd, R., Bowles, S., Camerer, C. F., Fehr, E., Gintis, H., … Tracer, D. (2005). "Economic man" in cross-cultural perspective: behavioral experiments in 15 small-scale societies. *Behavioral and Brain Sciences*, *28*(6), 795–855. http://doi.org/10.1017/S0140525X05000142

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *The Behavioral and Brain Sciences*, *33*(2–3), 61-83-135. http://doi.org/10.1017/S0140525X0999152X

Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., … Ziker, J. (2006). Costly punishment across human societies. *Science*, *312*(5781), 1767–70. http://doi.org/10.1126/science.1127333

Herrmann, B., Thoni, C., & Gächter, S. (2008). Antisocial Punishment Across Societies. *Science*, *319*(March), 1362–1367.

Horton, J. J., Rand, D. G., & Zeckhauser, R. J. (2011). The online laboratory: Conducting experiments in a real labor market. *Experimental Economics*, *14*(3), 399–425. http://doi.org/10.1007/s10683-011-9273-9

Iriberri, N., & Rey-Biel, P. (2013). Elicited beliefs and social information in

modified dictator games: What do dictators believe other dictators do? *Quantitative Economics*, *4*(3), 515–547. http://doi.org/10.3982/QE135

Lamba, S., & Mace, R. (2011). Demography and ecology drive variation in cooperation across human populations. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(35), 14426–14430. http://doi.org/10.1073/pnas.1105186108

Ledyard, J. (1995). Public goods: some experimental results. In *Handbook of experimental economics* (pp. 111–193). http://doi.org/10.1016/0037-7856(73)90129-7

Leung, A. K.-Y., & Cohen, D. (2011). Within- and between-culture variation: Individual differences and the cultural logics of honor, face, and dignity cultures. *Journal of Personality and Social Psychology*, *100*(3), 507–526. http://doi.org/10.1037/a0022151

Linquist, S. (2015). Which evolutionary model best explains the culture of honour? *Biology and Philosophy*, *31*(2), 1–23. http://doi.org/10.1007/s10539-015-9515-x

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), 426–441. http://doi.org/10.1037/0022-3514.57.3.426

Mackyn, P., Wilson, S., Fischetti, T., & Goworowska, J. (2011). Population Distribution and Change : 2000 to 2010. *U.S. Census Bureau*, *2010 Censu*(March), 1–12.

http://doi.org/http://www.census.gov/prod/cen2010/briefs/c2010br-01.pdf

McElreath, R. (2003). Reputation and the evolution of conflict. *Journal of Theoretical Biology*, *220*, 345–357. http://doi.org/10.1006/jtbi.2003.3166

Mengel, F. (2014). Risk and Temptation: A Meta-Study on Social Dilemma Games. *SSRN Electronic Journal*. http://doi.org/10.2139/ssrn.2519997

Nettle, D., Colleony, A., & Cockerill, M. (2011). Variation in Cooperative Behaviour within a Single City. *PLoS ONE*, *6*(10), e26922. http://doi.org/10.1371/journal.pone.0026922

Nikiforakis, N. (2008). Punishment and counter-punishment in public good games: Can we really govern ourselves? *Journal of Public Economics*, *92*, 91–112. http://doi.org/10.1016/j.jpubeco.2007.04.008

Nisbett, R. E. (1993). Violence and U.S. regional culture. *The American Psychologist*, *48*, 441–449. http://doi.org/10.1037/0003-066X.48.4.441

Nisbett, R. E., & Cohen, D. (1996). *Culture of Honor: The Psychology of Violence in the South*. Boulder, Colorado: Westview Press.

Nowak, A., Gelfand, M. J., Borkowski, W., Cohen, D., & Hernandez, I. (2015). The Evolutionary Basis of Honor Cultures. *Psychological Science*, *27*(1), 12–24. http://doi.org/10.1177/0956797615602860

Perugini, M., Gallucci, M., Presaghi, F., & Ercolani, A. P. (2003). The Personal Norm of Reciprocity. *European Journal of Personality*, *17*, 251–283.

Peysakhovich, A., Nowak, M. a, & Rand, D. G. (2014). Humans display a "cooperative phenotype" that is domain general and temporally stable.

Nature Communications, *5*, 4939. http://doi.org/10.1038/ncomms5939

Pitt-Rivers, J. (1965). Honour and Social Status. In J. G. Peristiany (Ed.), *Honour and Shame: The Values of Mediterranean Society* (pp. 19–77). London: Weidenfeld and Nicolson.

Reed, J. S. (1982). Below the Smith Wesson Line: Southern Violence. In *One South: An Ethnic Approach to Regional Culture* (pp. 139–153). Baton Rouge: Louisiana State University Press.

Rustagi, D., Engel, S., & Kosfeld, M. (2010). Conditional Cooperation and Costly Monitoring Explain Success in Forest Commons Management. *Science*, *330*(6006), 961–965.

Samuelson, P. A. (1954). The pure theory of public expenditure. *The Review of Economics and Statistics*, *36*(4), 387–389. http://doi.org/10.2307/1925895

Selten, R. (1978). The equity principle in economic behavior. In H. Gottinger & W. Leinfellner (Eds.), *Decision Theory and Social Ethics* (pp. 289–301). Dordrecht: Reidel.

Selten, R. (1991). Properties of a Measure of Predictive Success. *Mathematical Social Sciences*, *21*, 153–167.

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate ? *Behavioral and Brain Sciences*, *23*, 645–726.

Teyssier, S. (2012). Inequity and risk aversion in sequential public good games. *Public Choice*, *151*, 91–119. http://doi.org/10.1007/s11127-010-9735-1

van den Berg, P., Molleman, L., Junikka, J., Puurtinen, M., & Weissing, F. J. (2015). Human cooperation in groups: variation begets variation. *Scientific Reports*, *5*, 16144. http://doi.org/10.1038/srep16144

Yamagishi, T., Horita, Y., Mifune, N., Hashimoto, H., Li, Y., Shinada, M., … Simunovic, D. (2012). Rejection of unfair offers in the ultimatum game is no evidence of strong reciprocity. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(50), 20364–8. http://doi.org/10.1073/pnas.1212126109

Yamagishi, T., Mifune, N., Li, Y., Shinada, M., Hashimoto, H., Horita, Y., … Simunovic, D. (2013). Is behavioral pro-sociality game-specific? Pro-social preference and expectations of pro-sociality. *Organizational Behavior and Human Decision Processes*, *120*(2), 260–271. http://doi.org/10.1016/j.obhdp.2012.06.002